Analyzability and Semantic Associations in Referring Expressions

A Study in Comparative Lexicology

Analyzability and Semantic Associations in Referring Expressions

A Study in Comparative Lexicology

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. P. F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op woensdag 10 oktober 2012

klokke 11.15 uur

door

Matthias Urban

geboren te Filderstadt, Duitsland

in 1982

Promotiecommissie

Promotores:	Prof. dr. M. P. G. M. Mous Prof. dr. B. Comrie, Max Planck Institute for Evolutionary Anthropology, Leipzig
Overige leden:	Prof. dr. M. Koptjevskaja-Tamm, Stockholms Universitet Prof. dr. M. Vanhove, CNRS-LLACAN, Paris Dr. F. K. Ameka

La nature de l'homme est toute nature, omne animal*. Il n'y a rien qu'on ne rende naturel; il n'y a naturel qu'on ne fasse perdre.*

— Pascal, *Pensées*

Contents

Abbreviations and glosses

1	first person
2	second person
3	third person
ABL	ablative
ABS	absolutive
ABSTR	abstraction /abstract nominalizer
ACT	ACTIVE
ADJVZ	adjectivizer
AGT	agent (nominalizer)
AL	alienable
ALL	allative
AN	animate (class)
ANTIPASS	antipassive
APPL	applicative
ART	article
ASP	aspect
ASSOC	associative
ATT	attributive derivational suffix
AUG	augmentative
CAUS	causative
CISLOC	cislocative
CL / CLF / CLASS	classifier
COLL	collective (nominalizer)
COND	conditional
DAT	dative applicative verb prefix
DEC	decisive
DEF	definite
DEM	demonstrative
DERIV	derivational (semantically general derivational element)
DET	determiner
DIM	diminutive
DIR	directional suffix
DIST	distance
DN	dummy noun prefix
DUR	durative
EP/ EPEN	epenthetic
Eng.	English
ERG	ergative
FEM	feminine
Fr.	French
FREQ	frequentative
GEN	genitive
HAB	habitual

HR	adverbial directional suffix: hither, toward speaker
ILL	illative, inessive nominal case suffix
INAL	inalienable
INAN	inanimate
INDEF/ INDET	indefinite
INF	infinitive
INST.LONG	long-(shaped) instrumental prefix
INST/ MOUTH	instrumental causative prefix, with teeth/mouth
INSTR	instrument (nominalizer)
INTR	intransitive
IPFV	imperfective
IRR	irrealis complementizer
ITER	iterative
Lat.	Latin
LOC	locative (nominalizer)
M / MASC	masculine
MP	mediopassive
N / NEUT	neuter
NEG	negator
NMLZ / NR	nominalizer
NOM	nominal suffix
N.SFX	noun suffix
OBJ	unspecified object prefix
OBL	oblique relativizer
PART	particle
PASS	(medio)passive
PERF	perfective aspect
PFV	perfective
PST	past
P /PL	plural
POSS	possession/possessive
PROG	progressive
PTCPL	participle
QUAL	quality
RED	reduplication
REFLX	reflexive
REL	relativizer
RELAT	oblique relational suffix
REL.ABS	relativizing-absolutivizing suffix
RES	resultative
SBJ	subject
SCM	specific class marker
S / SG	singular
Span.	Spanish
SRFLX	semireflexive
STAT	stative
SUB	subordination
SUBJ	(nominal) subject
TOP	topic
THR	directional adverbial suffix: thither away from the speaker
VOL	volitive
WH	relativizing prefix

(NB: additional glosses for textual examples in chapters four and five are given in footnotes; in a few cases, glosses also appearing on this list have a different meaning there as stated in the respective footnotes).

Acknowledgments

It is a pleasure to acknowledge the nice people who accompanied me on the way to the completion of this work.

First and foremost, I am indebted to Bernard Comrie. Not only did he accept me as a PhD student, but also his advice helped to improve my work in uncountable ways. I am very grateful to him. This work would not exist without his support. At the same time, financial support by the Max Planck Society for the Advancement of Science is gratefully acknowledged. Also, I warmly thank Maarten Mous for accepting me as a buitenpromovendus at the Leiden University Centre for Linguistics and for his constant support.

I am greatly indebted to Johanna Mattissen and Sevim Genç for providing unpublished field data on Laz, to Angela Terrill, who made available unpublished lexical data on Lavukaleve, and to Mark Donohue, who provided unpublished Masur wordlists from which the Yei data used in this study are taken. I would also like to thank Tonya Stebbins and Julius Tayul for providing a pre-publication version of their Mali dictionary to me and Stuart Robinson for providing a version of his dissertation on Rotokas. My thanks are also due to Joseph Atoyebi (Yoruba), Ekaterina Gruzdeva (Nivkh), Andrej Nefedov (Ket), Pamela Munro with Catherine Willmond (Chickasaw), and Frank Seifart (Bora), who have checked data gathered from published sources for their respective languages of expertise and helped filling gaps in them. Without the kind support of all the abovementioned scholars, the empirical data on which this study is based would have been much more incomplete.

My sincere thanks are due to Bodo Winter, Michael Cysouw, and Roger Mundry for tirelessly answering my inquiries about the proper application of statistical methods on my data far beyond the call of duty. Hans-Jörg Bibiko equally tirelessly helped to improve the organization of the Filemaker database underlying this study and the visualization of patterns in the data by maps. Without their help, the interpretation of the data would have been much more cumbersome and time-consuming.

I would like to heartily thank the following linguists for their willingness to help me with literature written in languages that I do not read, for discussing certain aspects of the present work with me, or for answering questions of mine by e-mail: Dik Bakker, David Beck, Balthasar Bickel, Jürgen Bohnemeyer, Matthew Dryer, David Gil, Cliff Goddard, Mar-

tin Haspelmath, Iren Hartmann, Hagen Jung, Zaira Khalilova, Maria Koptjevskaja-Tamm, Ekkehard König, Peter Koch, Elena Lieven, Andrej Malchukov, Johanna Nichols, Sebastian Nordhoff, Carolyn O'Meara, Nick Riemer, Hans-Jürgen Sasse, Frank Seifart, Søren Wichmann, Alena Witzlack-Makarevich, and Anna Zalizniak.

Many thanks go out also to Moritz Deutschmann, Mark Whitten, Bodo Winter, Alena Witzlack-Makarevich, and Joshua Wilbur for proofreading parts of the dissertation, and to Eugénie Stapert for translating the summary of this work into Dutch. I am also grateful to her for acting as a paranimf at my defence, as I am to Rebecca Voll.

Finally, I would like to thank my friends at the Max Planck Institute for Evolutionary Anthropology, thanks to whom my time there was not only intellectually stimulating but also personally pleasant, and who shared the good and the bad with me. You know who you are.

None of the people mentioned above are to be held accountable for any mistakes and shortcomings in the coding, analysis and interpretation of the data, which are solely my responsibility.

Chapter 1

Introduction

The seminal moment to which this thesis owes its existence came in the summer of 2006 when I was a M.A. student in linguistics at the University of Cologne, long before I learned about the (far-flung) relevant literature and that there is something called "lexical typology." In the hall in front of the seminar room were old yellowed working papers of the institute from the seventies, lying there on a table waiting for someone to come by and take them away, which I did (in spite of suffering from a dust mite allergy, and these papers were *very* dusty). Among them were several that dealt with "descriptivity," by which the researchers understand a certain type of analyzability of lexemes. I then read Seiler (1975), in which this research program was introduced and explained. In this paper, I stumbled upon the following "descriptive" term for 'stone' from Cahuilla, a Uto-Aztecan language of California on which Seiler did fieldwork:

(1.) *qáw-iš*
harden-PFV (adapted from Seiler 1975: 24)

Cahuilla *qáwiš* is deverbal, formed by addition of a perfective aspect suffix to the verb stem: 'that which has hardened' (although, according to Seiler, it is conventionalized with the meaning 'stone' only nowadays). A couple of weeks later, I learned about the etymology of modern Germanic words for 'stone' completely by chance in a seminar. They can convincingly be shown to go back to the following Proto-Indo-European structure:

(2.) **stāi-no*
harden-PFV (adapted from Pfeifer 1993: 1353)

I was simultaneously intrigued by two facts: first, that languages actually do exist which have morphologically complex terms for such basic vocabulary items as 'stone,' and secondly, that two geographically, culturally, and temporally as remote languages as Cahuilla and Proto-Indo-European should have chosen to conceptualize the same meaning in precisely the same manner.

This anecdote serves to summarize the main concerns of the present thesis. First, it seeks to assess differences between languages in the number of morphologically complex lexical items. Scattered in the literature, one finds unsystematic statements about geographically and genetically widely dispersed particular languages and a felt preponderance for morphologically complex terms. At times, these statements make particular reference to the nominal domain, which will also be the focus of the present work (see chapter 3 for the reasons). Thus, Seiler (1976: 6) writes about Cahuilla (Uto-Aztecan):

> The analysability nd morphological transparency of a considerable portion of all nominal expressions ... is immediately recognizable.

Pawley (1993: 99), while mostly concerned with complex verbs, states that in Kalam (Trans-New-Guinea?),

> [t]here are no simple nouns for some conceptual categories which one might expect to be universal, such as those roughly translated by 'person, human being', 'parent', 'child', 'ancestor', 'enemy', 'rain', 'thirst'.

On Yélî Dnye, a language isolate spoken on Rossel Island, an island belonging to the Louisiade Archipelago located southeast of New Guinea, Levinson (2006b: 230) remarks that

> Yélî Dnye is a language where many important, commonly employed nominal concepts are expressed with compounds.

The clearest statement is made by O'Meara and Bohnemeyer (2008: 332-333) for Seri, an isolate of Mexico (sometimes affiliated with the putative Hokan family):

> Complex expressions ... are in fact pervasive in the Seri nominal lexicon. ... It is thus clear that paucity of monomorphemic lexicalization and compensatory use of complex descriptive terms is a general typological characteristic of the nominal lexicon of Seri.

More generally, Mithun (1999: 287), in a discussion of ritual speech registers, states that in North American languages

> [s]ingle words are often composed of many meaningful parts, and their literal meanings are in many cases still perceptible to speakers. Indeed, words in all domains are frequently coined from complex descriptions.

What these assorted statements show is that the differential degree of morphologically complex terms clearly is a typological variable that has not received much attention by present-day comparative linguists so far (though it is sketched prominently in Saussure 1916/1967, see chapter 2 for review), and they suggest that there are clear differences between languages here waiting to be systematized. This has not been done so far in a principled approach, in spite of a general awareness by typologists of these differences, as revealed by remarks such as Aikhenvald's (2007: 21), who observes that "[l]anguages differ in how much derivational motivation (and hence derivational complexity) they allow for individual words. ... Decomposable terms in some languages can correspond to non-decomposable ones in others." Therefore, the questions to be addressed in this work include: are there significant differences between the languages of the world with respect to the degree of morphologically complex terms in the lexicon, possibly correlating with the affiliation to a particular language family or a linguistic area? Is the predominance of simplex lexical items in the better-known European languages an "atypical phenomenon"

(Sasse 2001: 503), going back to extensive language contact and concomitant lexical borrowing? What, if any, is the role that the language-specific means of word-formation have to play? Is it appropriate to postulate a typological trait, as O'Meara and Bohnemeyer (2007) suggest for Seri, for languages with a pronouncedly high degree of complex formations in the lexicon also for other languages, and how, considering also the grammatical features and their interaction with the lexicon, could this trait be reasonably delimited? Likewise, and perhaps even more importantly, possible causes for these differences also essentially remain in the dark, and an important aspect of the present study will be to give reasons for the behavior of individual languages by searching for typological correlates, paying particular attention, as suggested by Dixon (2010: 257), to structural properties of the languages.

As to the second main aspect of the present work, differences and similarities in the semantic structures found in morphologically complex terms, some work is available, at least for individual meanings and semantic domains (see chapter 2 and dispersed references in chapter 6 for a more extensive discussion of the state of research and cross-references). However, the lexicon is vast, and there are clearly many more recurring patterns to be discovered. Parallels like those in (1.) are remarkable in so far as the terms and their underlying structure have most likely arisen independently of one another. Canart (1979: 66), noting the frequent semantic extension of 'skin' to 'fur,' 'feathers,' and 'scales,' says that "[w]hat is perhaps curious is that ordinary people all over the world discovered a number of these fundamental truths quite independently and in the most unrelated languages and cultures" (again, see chapter 2 for a more thorough review of such statements).

Note that Canart's statement about similarities in semantic extension pertains to terms not characterized by morphological complexity, but rather to monomorphemic simplex lexical items. Indeed, from the semantic point of view, semantic associations not realized by morphologically complex terms but by semantic extension are just as interesting. Under this aspect it is not fruitful and even artificial to keep semantic associations by morphologically complex terms and by polysemy strictly apart, as is argued in chapter 3, which will expose the minimal theoretical framework of the study. Following Koch and Marzo (2007), a useful cover term that is used in this study for ties between words and the meanings they convey, regardless whether they are realized formally by word-formation relations or by polysemous or ambiguous conflation in a single monomorphemic lexical item is LEXICAL MOTIVATION,[1] and the property of such terms consequently is that they are LEXICALLY MOTIVATED. Although it also has a prehistory briefly touched upon in chapter 2, motivation as a current term in linguistics goes back to Saussure (1916/1967), who also includes onomatopoeia under this umbrella term, and the qualifier "lexical" hence is to highlight the fact that this type of motivation is established not directly by sound symbolism, but by a mirroring of a semantic relation on the level of linguistic expression, that is, the lexical item.

Armed with this general concept allowing for investigation of motivated words of both types, there are questions one can ask about the semantic side of motivated lexical items: can the universal tendencies in the (cross-)linguistic realization of certain mean-

[1] As a convention, technical terms will be printed in small caps when first introduced throughout this work.

ings that are beginning to show on the horizon be consolidated and can they also be found for other meanings than those already investigated (see chapter 2 for review). How strong are the tendencies in each case? How do recurrent conceptualizations look like, and is it perhaps even possible to learn something from them about human cognition? Conversely, which patterns are rare, only found in a few languages? Are there, next to universal trends, also patterns that are peculiar to a certain area (and if so, what is their history?) or peculiar to certain grammatical properties (lexicon-grammar-interaction)? Generally, the spirit of the approach adopted is that of Matisoff (2004: 385):

> Are human thought processes as reflected in language everywhere the same? The extremes of relativism and universalism are equally to be avoided, in favor of an empirical approach that appreciates in equal measure the nuances of how languages differ and resemble one another.

Another perspective one can assume is to depart from a certain meaning and to ask what structure the terms designating it have cross-linguistically. How are certain semantic fields, such as meanings revolving around 'fire' and 'water,' organized cross-linguistically?

Matisoff (2004: 384-385) calls for a "massive international effort ... to create a master database of semantic associations in the world's languages." The present work surely cannot offer such an effort, but the results are initial steps in this direction.

Throughout, and concerning both main aspects just mentioned, the present work also seeks to shed light on the typological, historical, socio-cultural and, occasionally and very cautiously, also cognitive background of the similarities and also of the differences if they are found.

When attempting to situate the present work into the context of related typological research, it is clear that it is part of what has come to be called "lexical typology." This field has only in the past few years, with a workshop at the conference of the Association of Linguistic Typology in 2007 in Paris and publications such as Vanhove (2008), come to be recognized as a subfield of typological investigations of languages on its own (maybe, as Haspelmath 2003: 211 says, because "many linguists regard the study of grammar as more interesting and prestigious" when contrasted with the lexicon, perhaps an aftereffect of Bloomfield's 1933: 274 (in-)famous dictum that "[t]he lexicon is really an appendix of the grammar, a list of basic irregularities"). Luckily, the volume has received attention in the community, as evidenced by reviews such as Citarrella (2010), Newman (2010), and Traugott (2010), who in particular notes that previously available suggestions and hypothesis were based on either evidence from one language only or from a small set of languages. This thesis, having an explicitly cross-linguistic orientation, is another step towards remedying this situation. However, as with any young field of research, "lexical typology" is presently characterized by a multitude of coexisting methodologies which are based on varying background assumptions about the nature of meaning, the lexicon, and the relation between the two, with at times decidedly heterogeneous research goals (contrast the conceptions of Lehmann 1990 with that of Behrens and Sasse 1997, Koch 2001,

and with that in the work of Talmy, whose overview article on lexicalization patterns, Talmy 2007, notably bears the title "lexical typologies").

Given this orientational pluralism and the exploratory nature of the thesis, it is primarily data-driven, not theory-driven. This means that the theoretical framework to be developed in chapter 3 is intended to categorize formal and semantic relations typologically, but in a way that does not distort the data from a particular point of view (such as a particular theory of morphology, word-formation, or lexical semantics) and to allow the data to speak for themselves. It also means that the generalization and results to be presented in chapters 5 and 6 are to a large extent generated out of the data themselves, rather than through the application of general preconceived assumptions about the nature of "language" (in the sense of *langage*) or the cognitive infrastructure that renders it possible.

If lexicology is concerned with the study of the structure of the lexicon of a particular language, the topics dealt with here could be labeled for the time being and to use a term coined by Tappolet (1895: 2), as questions of COMPARATIVE LEXICOLOGY (in contrast to research programs concerned with semantics proper, such as Levinson and Meira 2003): the cross-linguistic comparative investigation of the structuring of the lexicon, both of the formal structure of the words it contains, the semantic fields they belong to, as well as of the ties these structures betray to other elements of the lexicon, and, in a second step, to ask why the lexicon is organized in this or that way in different languages and to motivate this behavior, be it on grounds of language-internal, areal-typological, or extra-linguistic reasons. Eventually, once the field of lexical typology is more consolidated, comparative lexicology might become one of its subbranches.

To be sure, this introductory discussion is merely meant as a brief panoramic vision of the topics to be discussed and to provide an attempt to roughly situate them in the context of existing research. As such, it does not do full justice to the copious amount of literature that exists on questions related to lexical motivation, and therefore, before setting out the framework of the present study in more detail and presenting its results, it is appropriate to delve further into the (pre-)history of the topic. This is undertaken in chapter 2. Readers not interested in such a discussion can skip this chapter and continue on to chapter 3 immediately without losing crucial information for the understanding of the framework presented there and hence for the rest of this book.

Chapter 2

Lexical Motivation Cross-linguistically: The History of an Idea

2.1. INTRODUCTION

The aim of this chapter is to provide an overview of the history of thoughts on the phenomenon of motivated terms, with the discussion being restricted to such works dealing with it in a comparative perspective. The reason for the inclusion of such a chapter is twofold: first, the topic has a long and rich history in European thought, being rooted in Renaissance Philosophy on the one hand and in the more practically oriented goal of effective language teaching on the other, the latter in the context of the evolving Humanist movement.

Second, a good portion of these publications is very far-flung, and there is no continuous tradition linking them together. It should be stressed that most authors writing on the topic do not seem to have been aware of their predecessors and therefore unknowingly were reinventing the wheel each time on their own. Thus, this chapter does not merely summarize an already established and well-known history of the topic that is easily accessible. One of its purposes is therefore to prevent some of the ingenious works on the topic from continuing to be buried in oblivion.

Some of the works to be discussed are quite old, and obviously do not adhere to the standards of modern research. What is the point in discussing all this old, nearly forgotten literature? Before embarking upon the purely linguistic discussion in the main body of the present work, it is worth noting that similarities and differences in the semantic structure of morphologically complex expressions have, either implicitly or explicitly, figured prominently, and were sometimes even at the heart of many of the trends and debates in the philosophy of language in the past three centuries. They have played a vital role in past approaches to the everlasting quest of working out the relation of what is culture-specific in human thinking to the grand scheme of the "psychic unity of man" (for an overview on this notion see Carneiro 2003: 17-18) and vice versa. In particular, many authors have assumed that the language-specific conceptualizations of individual concepts can be straightforwardly identified with the "völkergedanken" in the sense of Bastian (1870, 1881), and the corresponding concepts with the universal "elementargedanken" shared by all humans. It is this importance which has been assigned to the topic by many authors in the past that makes it, next to a great many other reasons which hopefully will become clear in the course of this work, a worthwhile enterprise to

reconsider it with the possibilities of today, although the big metaphysical questions connected with it in earlier times will not be dealt with any further.

The overview provided in the following cannot claim exhaustiveness, in particular as far as the older literature is concerned, since a detailed systematic exploration would have constituted a time-consuming research undertaking in itself. Nevertheless, it does incorporate an exemplary and representative selection of what may be taken to be the most relevant writers from earlier centuries on the topic. The aim of this chapter is to provide a short history of the idea of lexical motivation, not to develop the framework on which this study is based. Works that are both important contributions to the history of the idea and relevant for the development of the framework of the present study will be mentioned briefly here, and relevant aspects will be discussed in more detail in chapter 3.

2.2. EARLY CONSIDERATIONS

The discussion of this section is set roughly in the 16th century. Despite the huge interest in etymology in ancient Greece and Rome, acquaintance with multiple languages (other than Greek and Latin) was extremely restricted and therefore the empirical basis for a consideration of patterns of lexical motivation in a wider array of languages was largely lacking (note, however, that even in Plato's cratylus the "origin" of some words is sometimes sought in different languages or varieties of Greek). A similar situation held in the Middle Ages, where even the intellectual elite at universities was usually competent in their local vernacular language and Latin and even knowledge of Greek was highly unusual. This, however, does not exclude the possibility that relevant literature predating the 16th century may actually exist. Appraisal of the vernaculars and a sense of awareness of their diversity eventually arose in the Renaissance era, and it is precisely from this time that a variety of conceptions of language that recur to data from different languages is found, such as the etymological-philosophical-theological notion of the *harmonia linguarum*, which sought to reduce the diversity found in the (then known) languages of the world to a common *ursprache* of mankind, which was usually taken to be Hebrew. Also in this era, there are first traces of awareness for differences between motivational patterns in languages, both quantitatively and conceptually.

The first writer to be considered is the 16th century Spanish humanist and grammarian Francisco Sánchez de las Brozas, better known as Franciscus Sanctius Brocensis, the Latinized version of his name. Sanctius was, as many of his contemporaries, a strong believer in reason and rationality as the defining features of the human race, and therefore, since language is a product of the human mind, emphasized the logical structure and the reason enshrined in it (his work is often regarded as an important precursor of the Grammar of Port-Royal). For this reason, in the introduction to his then well-known and popular grammar of Latin entitled *Minerva o de causis linguae Latinae*, Sanctius sides with Plato's Cratylus in the question of the nature of the linguistic sign (arbitrary vs. motivated), since he cannot imagine that the rational human mind should create a mental lexicon that is characterized by methodless and haphazard form-meaning associations. Instead, he takes the position that one true, non-arbitrary etymology for each expression must exist

which mirrors its rational structure. It is in this context that he discusses the following possible objection:

> Sed dices: qui potest fieri ut uera sit nominis etymologia, si una eademque res uariis nominibus per orbem terrarum appellatur? Dico: eiusdem rei diuersas esse causas, quarum illi hanc, nos aliam contemplamur. Sic Graeci ἄνεμον, Latini uentum appellauere: illi a spirando, hi a ueniendo. Fenestram a φαίνεσθαι deduxit Latinus, "ventana" a nostris dicitur, Lusitanis "ianella," quasi parua ianua.
>
> But you will say: How can it be possible that there is a true etymology of a name, if one and the same thing is designated by various names throughout the globe? I say: the same thing has diverse causes, of which they point their attention to this one, we to another one. Thus, the Greeks call the wind *anemos*, the Romans *ventus*, the former from blowing [*anemidzein*], the latter from coming [*venire*]. The Romans derived *fenestra* ['window'] from *phainesthai* ['to appear'], by our people it is called *ventana* [showing a connection to the aforementioned *ventus*] and by the Lusitanians[1] *ianella,* as if it were a little door [*ianua* 'door' + diminutive suffix]. (1995/1587: 40, translation and additions in square brackets by the present author, which is always the case also for following translations unless otherwise indicated).

What sets Sanctius apart from most of his contemporaries is that he actually cites real data from a number of languages in his discussion of the matter (albeit often drawing the wrong conclusions) instead of excogitating hypothetical connections between words in just one language (see also Percival 1988 for historical discussion).

Another humanist writer who displayed a remarkable awareness for the issue is the Czech Johann Amos Comenius (writing about 80 years after Sanctius) who is today remembered primarily for his pedagogical and educational writings. Lexical Motivation is discussed in various passages of his *Novissima Linguarum Methodus*, a lesser known work of his dealing with translation. Comenius commanded a large number of languages from different genera of Indo-European as well as Hebrew, and thus it is not surprising that it occurred to him that it is sometimes impossible to render a morphologically complex word in one language with an equally complex term in another, or a morphologically simplex word with an equally monomorphemic one.

> Evitari tamen, quin se Primigeniarum exercitui derivatæ quædam et compositæ admiscerent, non potuit. *Primùm,* quia vernaculis quibusdam primitivis, quæ responderent Latinæ primitive, defuerunt: ut *schlecht, Woche, Mangel, Üben* etc., quae Germanis primitiva sunt, Latinè autem nonnisi per derivatas et compositas illas *Simplex, Septimana, Defectus, Exercere* etc. reddi possunt (2005/1648: 295-296).
>
> Still, he [the translator in the process of translating from Latin into modern vernacular languages] could not avoid a certain amount of scrutinized derived words and compounds mixing themselves in. First, because the vernaculars may lack an amount of simple words

[1] That is, the Portuguese.

> which correspond to simple words in Latin, like *schlecht, Woche, Mangel, üben* ['simple,'[2] 'week,' 'lack,' 'practice'] etc., which are simple in German, but can nevertheless be expressed by derivatives and compounds in Latin, like *simplex, septimana, defectus, exercere*.

In addition, Comenius was also sensitive to differences in Word-Formation techniques:

> Hoc item in Vocibus observandum est, quód quævis lingua *primigeniarum* habet numerum quendam certum, haud ita multum, à quibus deducunt caeteras. Differunt autem, quòd quædam solâ *derivatione*, ut ferè *Hebræa* cum cognatis suis; quædam sola *compositione*, ut ferè *Germanica*; quædam utroque modo, ut Græca, Latina, Slavonica (2005/1648: 74).
>
> It can be observed that any given language has a certain number of principal words, often not so many, from which it derives the others. Languages differ, however, because certain ones do so only by derivation, like in general Hebrew and its cognates, certain ones only by composition, like in general German, and others by both means, like Greek, Latin, and Slavonic.

Comenius also had a very fine sense of similarities in the conceptual content in terms with the same meaning across languages. Thus, with unmistakable delight and great sarcasm, he ridicules a certain Georg Phillip Harsdörffer, a German, who had proposed in 1646 that the Polish and Hungarian term *weiwod* 'duke' is etymologically identical with the Celtic term *witdod* 'philosopher' (itself a loan translation of Greek *philosophos*). Comenius comments on this by saying that it is certainly not a mistake for a translator or anyone who deals with language in general to be multilingual, since, had Harsdörffer known any Polish, it would have occurred to him that *weiwod* is in fact identical in conceptual structure to the corresponding term in his native language, *Herzog,* and he would not have proposed his venturesome etymology (2005/1648: 463, both terms were in fact, as correctly pointed out by Comenius, originally compounds of the respective terms for 'army' and 'leader').

A third early work on the subject that is relevant in the present context and, as will be seen, the ideas of which were an inspiration to certain linguists, is Giambattista Vico's *Scientia Nuova*, first published in 1725 and with a lot of reworking as a second and final edition in 1744. Vico's philosophical program may be summarized in a nutshell as the quest for the (hidden) underlying unity of humankind in all the diversity in conduct and customs of the different peoples of the world. Inspired by the universal language schemes which had been developed in a bewildering array of varieties by 17th century writers such as Wilkins, Dalgando and Leibniz (see Cram and Maat 2000), one place where Vico was looking for this unity was, of course, language. Vico believed that varying living conditions had superimposed regional peculiarities to the underlying sameness of the peoples of the world and their culture, which according to him are mirrored in language:

> ... [C]ome certamente i popoli per la diversità de' climi han sortito varie diverse nature, onde sono usciti tanti costumi diversi; così dalle loro diverse nature e costumi sono nate

[2] In present-day German, *schlecht* primarily means 'bad.' At the time Comenius was writing, the dominant meaning was however 'plain, simple, frugal,' a meaning associated in Modern German exclusively with the variant form *schlicht* (Kluge 2002, s.v. schlecht, schlicht).

> altrettante diverse lingue: talchè, per la medesima diversità delle loro nature, siccome han guardato le stesse utilità o necessità della vita umana con aspetti diversi, onde sono uscite tante per lo più diverse ed alle volte tra lor contrarie costumanze di nazioni; così e non altrimenti son uscite in tante lingue, quant'esse sono, diverse. (1976/1744: 421-422)

> [A]s the peoples have certainly by diversity of climates acquired different natures, from which have sprung as many different customs, so from their different natures and customs as many different languages have arisen. For by virtue of the aforesaid diversity of their natures they have regarded the same utilities or necessities of human life from different points of view, and there have thus arisen so many national customs, for the most part differing from one another and at times contrary to one another; so and not otherwise there have arisen as many different languages as there are nations (Bergin and Fish 1984: 148).

The further discussion makes clear that by "different points of view," Vico, assuming a quasi-onomasiological perspective, refers to the different ways in which the same concept may be verbalized in different languages:

> Perciò da noi in quest' opera la prima volta stampata si è meditata un' *Idea d'un dizionario mentale da dare le significazioni a tutte le lingue articulate diverse*, riducendole tutte a certe unità d'idee in sostanza, che, con varie modificazioni guardate da' populi, hanno da quelli avuto vari diversi vocaboli... (1976/1744: 422)

> And for this reason we excogitated, in the first edition of this work ..., an Idea of a Mental Dictionary for assigning meanings to all the different articulate languages, reducing them all to certain unities of ideas in substance, which, considered from various points of view, have come to be expressed by different words in each (Bergin and Fish 1984: 148).

Vico mentions proverbs which, although they express the same semantic content, are coined from varying notional templates in different languages (such as, for instance, English *to have one's cake and eat it too* vs. German *auf zwei Hochzeiten tanzen*, literally 'to dance on two weddings') as one of several examples to add flesh to these theoretical considerations. Unfortunately, there are no lexical examples taken directly from actual languages, but Vico's discussion makes it quite clear that his thoughts are applicable to languages' lexical structure as well. Vico's work is important, because, although deeply rooted in contemporary thinking[3], he is not concerned with pseudo-etymological guesswork but

[3] Vico himself cites Hayne (1639) as an inspiration and as a brother in mind, although Hayne is concerned more with the rather traditional topic of the then popular concept of the *Harmonia Linguarum* (see also Percival 1988 for discussion). The only relevant passage from Hayne that Vico may refer to runs as follows:

> Qui [Hayne's colleague Henry Jacob, MU] mihi inter colloquendi familiaritatem saepius insinuavit adeo se in verborum causas penetrasse, ut Artis quoddam Etymologicae Systema sibi fabricavit, & cum docti priores universi rem Etymologicam, aut carptim quidem atque obiter tractarint, aut cum plenissime, non nisi voces ordine Alphabetico Primitivas disposuerint; ipse profitetur se methodo sua & hactenus intentata, per certas proprietatum classes subordinatim velle istas disponere, donec in paucissimis desinant Principiis: ut haec verborum Philosophia exacte imitetur illam rerum. ... Sic tandem (quod primum mihi optandum videbatur) evadet haec Harmonia Linguarum seu Ars Etymologica multo perfectior, & linguarum studiosis fructuosior (Hayne 1639: 64-5).

rather assumes what one may call a synchronic orientation. Even more importantly, the informal notion of "different points of view" enshrined in linguistic expressions in different languages for the same extra-linguistic referent, implicit for instance in Sanctius's writing, is fully articulated for the first time in Vico, and, as will emerge from the ensuing discussion, it will continue to turn up at various points in later centuries, up to contemporary linguistic research.

2.3. HUMBOLDT AND AFTERMATH

The next author that will be shown to have discussed the idea of "different points of view" in languages' expressions is probably the most famous and the most influential to be discussed here: Wilhelm von Humboldt. It is of course not the point of this discussion to attempt to deal with the enormous exegetical literature on Humboldt, but rather to demonstrate, using various passages of his writings, that considerations very similar to the ones already seen in earlier writers figure prominently in important passages of Humboldt's work. For instance, in the discussion of the famous (or perhaps infamous) concept of *innere sprachform*, often considered the central notion of his work, Humboldt makes it very clear that an important part of this idea consists of the observation of different aspects highlighted by terms for one and the same object in different languages or even within one language. Humboldt's example is the 'elephant' in classical Sanskrit, which features a variety of terms for this animal that are motivated by different aspects of its appearance and behavior:

> Wie bei der Lautform als die beiden hauptsächlichsten zu beachtenden Punkten die Bezeichnung der Begriffe und die Gesetze der Redefügung erschienen, ebenso ist es in *dem inneren, intellectuellen Theil* der Sprache. ... Denn es muß innerlich jeder Begriff *an ihm selbst eigenen Merkmalen, oder an Beziehungen auf andere* festgehalten werden, indem der Articulationssinn die bezeichnenden Laute auffindet. Dies ist selbst bei äußeren, körperlichen, geradezu durch die Sinne wahrnehmbaren Gegenständen der Fall. Auch bei ihnen ist das Wort nicht das Äquivalent des den Sinnen vorschwebenden Gegenstandes, sondern der Auffassung desselben durch die Spracherzeugung im bestimmten Augenblicke der Worterfindung. Es ist dies eine vorzügliche Quelle der Vielfachheit von Ausdrücken für die nämlichen Gegenstände; und wenn z.B, im Sanskrit der Elephant bald der zweimal Trinkende, bald der Zweizahnige, bald der mit einer Hand Versehene heißt[4], so sind

---

> He [Hayne's colleague Henry Jacob] repeatedly communicated to me in intimate interlocutions that he had penetrated so deep into the causes of words that he had fabricated himself a certain etymological system; and whereas all the earlier scholars treat the issue of etymology either somewhat sporadically and haphazardly or, if they treat it in any detail, still merely arrange the simple words in alphabetical order, he teaches that he wants to arrange them, following his so far unattempted method, according to certain sets of distinctive features until they end up resting on very few principles, so that the philosophy of the words precisely imitates the one of the things [which they designate]. ... Thus, finally, (what seemed primarily desirable to me) the Harmony of Languages or Art of Etymology would emerge much more perfectly and fruitfully for the language student.

[4] The explanation for the "twice-drinking one" is that the elephant first imbibes water with its trunk and only then pours it into its mouth. The explanation for "the one equipped with a single hand" is a metaphorical transfer of the human hand to the elephant's trunk.

> dadurch, wenn auch immer derselbe Gegenstand gemeint ist, ebenso viele verschiedene Begriffe bezeichnet. Denn die Sprache stellt niemals die Gegenstände, sondern immer die durch den Geist in der Spracherzeugung selbstthätig von ihnen gebildeten Begriffe dar; und von dieser Bildung, insofern sie als ganz innerlich, gleichsam dem Articulationssinn vorausgehend angesehen werden muß, ist hier die Rede. (Humboldt 1998/1832: 213-14, emphasis added)

> Just as designation of concepts and the laws of syntax appeared, in the sound-form, as the two points chiefly to be noted, so the same holds good in the *inner, intellectual part of language*. ... For every concept must inwardly be held fast *to markers peculiar to itself, or to relations with other concepts*, while the sense of articulation discovers the designating sounds. This is even the case with external physical objects that are plainly perceivable by the senses. Even for them the word is not the equivalent of the object that hovers before the sense, but rather the conception thereof through language-production at the particular moment of finding the word. This is a notable source of the multiplicity of expressions for the same objects; and if in Sanscrit, for example, the elephant is now called the twice-drinking one, now the two-toothed one, and now the one equipped with a single hand, as many different concepts are thereby designated, though always the same object is meant. For language never represents the objects, but always the concepts that the mind has spontaneously formed from them in producing language; and this is the forming under discussion here, insofar as it must be seen as quite internal, preceding, as it were, the sense of articulation. (Losonsky 1999: 83-84, emphasis added and original emphases removed).

It is very clear that, at least in the context of this example, the "inner, intellectual part of language" is the conceptual content of the individual motivated terms (in this case, the Sanskrit terms for 'elephant'), and its source concepts are referred to by the phrase "to markers peculiar to itself, or to relations with other concepts." From these and other observations, Humboldt concluded that there is a mental peculiarity of the speech community enshrined in its language which corresponds closely to the preferences a language makes in the semantic connections in its lexicon (in Humboldtian terms, the "Nebenideen" or "subsidiary ideas" of words, as represented by the reference to certain aspects of the entity to be named) and in those areas of its grammar which are related to the conceptual organization of the environment.

For example, a few pages later Humboldt mentions the fact that in a North American language with an animate/inanimate-distinction in its pronoun system, the stars are treated as animate grammatically and that the stars must therefore be conceived of as human-like by speakers of this language (an assumption which is, of course, problematic). A similar spirit can be found in the following lines of an earlier work of Humboldt written in French (interestingly, on the languages of North America as well), which is worth quoting here because a very concise phrasing of the above ideas is found in it and, more importantly, a possible comparative perspective is outlined, using the semantic domain of the human intellectual faculties as an example:

> On a souvent observé que les termes qui servent dans différentes langues à exprimer les mêmes objèts, surtout s'il s'agit d'idées ou de sentimens, diffèrent beaucoup dans les nuances plus fines de leurs acceptions. En analysant exactement chacun de ses termes, en se déterminant avec précision la valeur, et en les comparant ensuit ensemble, on acquiert une idée beaucoup plus parfaite et plus complette de l'objèt même qu'ils dénotent. Chaque

> mot présentant une idée nuancée d'une certaine manière, et ces nuances provenant d'un côté de la nature de l'objèt, de l'autre de la façon de le saisir, on apprend à connoître l'un et l'autre dès qu'on s'élève à un point de comparaison général; tandisque le raisonnement purement abstrait ne conduiroit jamais qu'imparfaitement à établir ces nuances, et par conséquent à embrasser toute l'étendue et toutes les modifications de l'objèt. On pourroit de cette manière faire un travail aussi utile que piquant sur les Synonymes dans différentes langues. ... En analysant et en comparant p.e. les mots qui dans les langues savantes de l'Antiquité, et nos modernes les plus cultivées désignent les facultés intellectuelles de l'homme, on feroit un cours pratique de cette partie de la Psychologie d'autant plus intéressant qu'on y découvriroit la manière de penser et de sentir de nations entières (Humboldt 1963/1821: 314-315)

> One has frequently observed that in different languages the terms that serve to express the same objects, especially when one is dealing with terms for ideas or sentiments, differ very much in highly subtle semantic nuances. In analyzing closely each of these terms, in determining with precision the content, one acquires a highly consummate and complete idea of the very object they designate. Every word presents a nuanced idea in a certain manner, and these nuances stem from an aspect of the nature of the object, and on the other hand from the manner of conception, and as one becomes acquainted with one and the other, one elevates from there to a point of general comparison; whereas pure abstract reasoning would never lead to establishing its nuances other than imperfectly, and consequently to embracing the whole variation and modifications of the object. In this manner one could produce a work as useful as acute on the synonyms in different languages. ... In analyzing and comparing for instance the words which in the savage languages of antiquity and in our modern, more cultivated languages designate the intellectual capabilities of man, one could produce a practical course of this part of psychology which is all the more interesting in that one would discover the mode of thought and sensation of entire nations. (translation with help from Nadège Lechevrel)

2.4. VÖLKERPSYCHOLOGIE

There is a direct connection between both Vico's and Humboldt's ideas in research within the *völkerpsychologie* paradigm, which can be roughly described as an amalgam of Humboldtian ideas with influences from ethnology, the emerging discipline of psychology, and earlier work with respect to the issue at hand. In 1869, a review (Eberty 1869) of a book on Vico by Carlo Cantoni was published in the accompanying journal to the research program, edited by the leading völkerpsychologists Steinthal and Lazarus and entitled *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*. Eberty (1869: 453) interprets the above quoted passages from Vico to the effect that "[Vico] glaubte, man könnte ein Universal-Etymologikon machen, welches nach der Wortbezeichnung darstellte, wie dieselbe Sache von den verschiedenen Völkern angeschaut ward" / "Vico believed one could create a universal etymologicon which displays according to the denomination how the same thing was beheld by the different peoples." This undertaking is enthusiastically welcomed by August Friedrich Pott (1974/1884-1890]: 42), who had been closely associated with the völkerpsychologie movement. But Pott also published an interesting and unfortunately forgotten article on the subject two decades earlier in the aforementioned journal entitled *Über Mannichfaltigkeit des sprachlichen Ausdrucks nach Laut und Begriff* (Pott 1860). In this, he essentially takes up the "different points of view"-idea already familiar from Vico, in

which the fundamental unity of mankind manifests itself. Like Vico, Pott believes in "die Einheit des reinen Gedankens" / "the unity of pure thought," which is the same among all humans when contemplating a certain entity in the extra-linguistic world, but which is expressed to by a "oft gar bunt ausſehender sprachlicher Ausdruck für denſelben" / "often quite colorful looking linguistic expression for it." For Pott, these linguistic realizations are symptoms of different "Volks-Logiken" / "folk-logics" (1860: 254) which manifest themselves in the different individual languages (note again the close connection to Humboldt) by highlighting certain aspects of the referent while at the same time necessarily neglecting others. The following quotes illustrate this line of thought:

> Es kann aber auf der anderen Seite eben ſo wenig befremden, wenn, namentlich grundverſchiedene Sprachen, zu Bezeichnung d e ſ ſ e l b e n Objects oftmals nicht bloß zu ganz verſchiedenen L a u t e n, ſondern auch zu sehr abweichenden B e g r i f f s - Vermittelungen, greifen, d.h., anders ausgedrückt, zu S y n o n y m i e n von innerlich mitunter aufs äußerſte unter ſich disparatem e t y m o lo g i ſ c h e n Werthe, welche - deſſenungeachtet- in ihrem Gegenstande, wo nicht ſich decken, doch als in ihrem gemeinſchaftlichen Z i e l p u n k t e zuſammentreffen müssen (Pott 1860: 256).

> On the other hand, it can be of equally little surprise if entirely different languages resort to not only completely different s o u n d s, but also to very differing concept-arrangements to designate the same object, i.e., in other words, to s y n on y m i e s of internally at times extremely disparate e t y m o l o g i c a l value, which, in spite of that, if they do not coincide, have to converge with regard to their subject matter in their common goal.

> [A]lle Benennung von Subſtanzen geſchieht immer nur fragmentariſch, gleichſam als müßte ein Bruchtheil (ein Merkmal, Epithet) ſtellvertretend die alleinige Pathenſchaft für das Ganze (die Subſtanz als Inbegriff einer Vielheit von Merkmalen) übernehmen und ihm ſ e i n e n Namen leihen. Die große Mannichfaltigkeit von Namen, die ein Ding haben kann, rührt eben daher, daß man bei der Benennung bald auf dieses bald auf jenes Merkmal (es hat aber deren eine große Menge) ſein Augenmerk richten kann ... (Pott 1860: 345).

> Every designation of substances always occurs solely in a fragmentary way, as if a fraction (a characteristic, epithet) had to take over representatively the godfatherhood for the whole (the substance as epitome of a multiplicity of characteristics) and lend it i t s name. The great multifariousness of the names a thing can have even stems from the fact that one can turn one's attention in designating either to this or to that characteristic (of which there is a large number).

Pott goes on to give a number of examples for a selection of concepts from different semantic domains for what he means, using data from a wide range of languages, given the time in which he is writing. For instance, he cites several equivalents for the concept 'bat,' which is expressed in some languages with reference to the fact that these animals usually begin their activities in the evening (such as Latin *vespertilio*, which contains *vesper* and Danish *aftenbakke*, which contains *aften*, both meaning 'evening'), whereas other languages employ "Benennungen, worin das Flattern ... hervorgehoben wird" / "denominations wherein the fluttering is highlighted," such as German *fledermaus* and Dialectal Spanish

raton volante, literally 'flying mouse,' and still others make reference to the animals' leathery wings as opposed to birds' feathers, like Sanskrit *aǵinapattra* ("Hautflügler") and Hungarian *bör-eger*, literally 'leather mouse.' Setting an even broader agenda, Pott outlines a research program consisting of gathering similar data for more concepts from a wide array of languages in order to thus establish something like a natural ontology of how humans perceive the world that surrounds them, as represented by their languages:

> Nicht minder würde eine Sammlung der von einem N a t u r g e g e n ſ t a n d e in den verſchiedenen Sprachen üblichen Benennungen gewiſſermaßen den Dienst leiſten einer Art Naturbeſchreibung deſſelben, die, wenngleich nicht auf wiſſenschaftlicher, doch im Uebrigen oft auf äußerſt ſcharfer und naturgetreuer Beobachtung ruht. Nur müßte man ſie erſt aus allen Winkeln und von aller Welt Enden her zuſammenleſen, weil in jedem einzelnen Namen dem Gegenstande doch nur e i n e, wenn auch an ſich ſehr hervorſpringende und charaktervolle Seite abgewonnen worden (Pott 1860: 345).
>
> Just as well a collection of the conventional designations for a natural object in the different languages would to a certain extent render the service of a sort of a natural description on it which rests, albeit not based on scientific observation, but nonetheless on acutely sharp and lifelike observation. One would only first have to glean it from all corners and all ends of the world, because in every single name only one, albeit an in itself very salient and characteristic aspect is wrested from the object.

Of course, at the time Pott wrote access to information about remote languages was extremely restricted, and today one encounters a much better, albeit not ideal, situation to assemble data "von aller Welt Enden her" / "from all ends of the world," which makes Pott's achievements even more remarkable.

Interestingly, Pott and other völkerpsychologists (like e.g. Lazarus and Steinthal 1860) also explicitly refer to the Humboldtian notion of *innere sprachform*, which Pott summarizes as "die der Benennung zugrunde liegende concrete und partielle Anſchauung" / "the concrete and partial conception which underlies denomination" (1860: 358), thus situating it exclusively in the domain of the lexicon and not in the grammar. Pott presupposes *innere sprachform* to be known to the reader in this interpretation.[5] This adoption of the (partial) identification of *innere sprachform* with the "Anschauung" putatively enclosed in a language's lexicon is also present in the work of the founding father of the völkerpsychologie movement, Wilhelm Wundt. Wundt developed a dichotomy between "concrete" and "abstract" thinking, which is conceived of by Wundt as one dimension of *innere sprachform* and can, according to him, be established by investigating the lexical characteristics of languages of different peoples. The former type of thinking is, in his opinion, found in so-called primitive societies, the latter in the more advanced languages of the "civilized" western societies (Wundt was one of the last defenders of the view that languages develop step by step to a more perfect level, the most advanced level

[5] Probably the last thing linguistics needs is yet another opinion on *innere sprachform*, but the interpretations by earlier scholars as well as certain passages in Humboldt's work themselves point in the direction that the same notion was at the core of the original conception of *innere sprachform* as well.

being represented by the Indo-European languages). Among the characteristics associated by Wundt with "concrete thinking" are the following:

> Die Erscheinungen sind darum von doppelter Art: sie bestehen erstens in dem Mangel zusammenfassender Bezeichnungen für verwandte Vorstellungen, und zweitens in der Substitution bestimmter Einzelvorstellungen, denen irgend ein allgemeiner Begriff als Merkmal zukommt, für diesen Begriff selbst. Konkrete Ausdrucksweisen der ersten Art sind es z.B., wenn eine Sprache den Menschen nicht als allgemeinen Gattungsbegriff, sondern nur in seinen besonderen Arten, als Mann, Weib, Kind u. dgl. kennt; solche der zweiten Art, wenn sie die Zahl 'vier' durch 'Zehen des Straußes', 'fünf' durch 'Hand', 'zwanzig' durch 'Mensch' bezeichnet (Wundt 1904: 443-444).

> The phenomena are therefore of a double nature: they first consist in a lack of abstracting designations for related conceptions, and second in the substitution of certain individual conceptions, which are characterized by some general notion, for this conception itself. Concrete terminologies of the first kind are if for instance a language does not know the human being as a generic term, but only in its particular kinds, such as man, woman, child, and the like; those of the second kind, if it designates the number 'four' by 'toes of the ostrich,'[6] 'five' by 'hand,' 'twenty' by 'human being.'

In effect, Wundt assumed, first, that when terms for which a conceptual source different from their own meaning can be indicated, this source plays a fundamental role in the constitution of the meaning (the ways in which it is "thought of"). This problematic equation of literal meanings, as found in languages of different peoples, with thought is, however, not at all a new nuance of the idea in Wundt. As seen earlier, such an equation is implicitly made as well by most earlier writers that were discussed, and it can be traced through the centuries along with the idea itself. Second, according to Wundt, from the presence of many such cases in a given language one can infer that the people speaking it have not yet made the transition to the allegedly more advanced modes of thinking as found in the more "abstract" European languages. This problematic idea will not be dealt with any further; it is sufficient to note that lexical data from languages which have something to do with the topic of this work play a crucial role in Wundt's theorizing. Indeed, the discussion so far has shown that the bold conclusion that the aspects highlighted by the (different) conceptualizations in complex expressions can reveal something about the way speakers think have left traces through centuries of reasoning on language and its relation to culture and cognition (see chapter 6 for some more cautious thought on what evidence is needed for such claims on the basis of the data collected for this study).

---

[6] Wundt does not mention the language in which this etymology is found, nor does he quote a source for this example. The source is likely Dobrizhoffer (1822: 169), who states that the South American language Abipón (Guaicuruan family) has only numerals up to three and that the speakers "make up for the other numbers by various arts: thus, *Geyenk ñatè*, the fingers of an emu, which, as it has three in front and one turned back, are four, serves to express that number." Dobrizhoffer is quoted in Tylor (1871), a widely read book in the 19th century, and it is likely that this example made its way into Wundt's work via Taylor. Incidentally, there is no mention of such a denomination in the dictionary part of Najlis (1966), where Abipón is said to have two stems for the numeral four neither of which resembles remotely the term mentioned by Dobrizhoffer.

In a different context, namely in a more general discussion of principles according to which objects are named, Wundt presents different examples for differing conceptualization strategies for the same concept similar to the one found e.g. in Pott (1860):

> So ist die *Erde* dem Römer die *trockene*, wohl im Gegensatz zum Meere (*terra* = **tersa* verwandt mit *torrere* dörren), dem Griechen die fruchtbare (*γῆ, γαῖα*, vielleicht verandt mit *γύα* Saatfeld), dem Germanen die *bewohnte* oder *bebaute* (ahd. *ërda*, wohl zusammenh. mit *artôn*, bewohnen, bebauen, lat. *arare*) (1904: 498).

> Thus the *earth* is for the Roman the *dry one*, probably as opposed to the sea (*terra* = * *tersa*, cognate with *torrere* 'desiccate'), for the Greek *the fertile* (*γῆ, γαῖα*, perhaps cognate with *γύα* 'seed plot'), for the Teuton it is *the inhabited or cultivated* (Old High German *ërda*, arguably associated with *artôn*, 'inhabit, cultivate,' lat. *arare*)[7]

In addition to merely presenting data from different languages, Wundt also develops a psychological account of the phenomena recurring to the notion of apperception (first used by Leibniz), which according to him governs the naming process by selecting a salient feature of the object to be named which is then fused in language with the conception of the object itself:

> In jedem Fall bezeichnet also das Wort eine zusammengesetzte Vorstellung, innerhalb deren ein Bestandteil im Augenblick der Benennung als der dominierende apperzepiert wurde (1904: 499).

> Thus in any case the word designates a composite conception, within which a component part has been apprehended in the moment of designation as the dominant one.

Wundt calls the feature of the object selected for naming the dominating feature ("dominierendes Merkmal"). Importantly, there is, in Wundt's account, a fusion of the semantics of the dominating feature selected in the process of apperception, and the object to be named.

Wundt's (and his predecessors') view found a profound critic in Anton Marty, a pupil of Franz Brentano, writing about three decades later. Marty criticized, often for very good reasons, almost every aspect of Wundt's work. Although he adheres to the notion of *innere sprachform* in principle as further developed by the völkerpsychologists as well, he has a very different idea as to what amount of importance should be assigned to the source concepts used to conceptualize meanings in Wundt's "concrete thinking" and the role they can play in determining the psychological characteristics of the people using them. Marty effectively denied that a position with respect to object naming such as Wundt's, in which there is no clear semantic distinction between the conceptualization process and the semantics of the term, is defendable at all:

[7] While the connection between Latin *terra* and *torrere* is corroborated by modern research (Wodtko et al. 2008: 701), the Greek and German etymologies proposed by Wundt seem uncertain (for the latter see Kluge 2002, s.v. Erde).

> Vor allem gehört die Vorstellung 'Zehen des Straußes', 'Hand' usw. hier nicht zum Inhalt des durch die Sprache ausgedrückten Denkens, wie ich schon anderwärts ausführlich dargetan. Inhalt ist der Begriff 'vier', 'fünf' usw. ebensogut wie wenn wir die Namen 'vier', 'fünf' usw. gebrauchen. Damit ist aber nicht gesagt, daß 'Zehen des Straußes' etwa eine Denkform für diesen Inhalt sei. Es gehört vielmehr überhaupt nicht zum Gedachten im Sinne der Bedeutung; die Vorstellung 'Zehen des Straußes' ist eine zur Vermittlung des Verständnisses dienende Begleitvorstellung ... also ein Stück Ausdrucksmittel, nicht zur Bedeutung gehörig, weder als Form noch als Inhalt. Wenn man sich manchmal mit Rücksicht darauf, daß diese für die gleiche Bedeutung des Verständnis vermittelnder Bilder in verschiedenen Sprachen verschieden sind, so ausdrückt, daß man sagt, die eine Sprache fasse den Begriff so, die andere anders auf, oder die eine denke in dieser, die andere in anderer Form, so kann mit diesen verschiedenen 'Denkformen' eben nur ein verschieden sprachliches, d.h. zu den Mitteln der Verständigung gehöriges Denken gemeint sein, nicht ein solches, das irgendwie die Bedeutung bildete (1950: 62).

> First of all, the conception 'toes of the ostrich,' 'hand' etc. does not belong to the content of thought as expressed by language, as I already demonstrated at length elsewhere. The content is the concept 'four,' 'five,' etc. just as if we use the names 'four,' 'five,' etc. This does not however entail that 'toes of the ostrich' is a form of thought for this content. In fact it does not belong at all to that which is thought in the semantic sense; the conception 'toes of the ostrich' is a subsidiary conception that serves to mediate understanding ..., a piece of expressive means, not belonging to semantics, neither as form nor as content. If one sometimes articulates oneself with respect to the fact that these images mediating understanding are different in different languages to the effect that one languages conceives of the notion in this way, another in a different way, or one language thinks in this, another in a different form, then what can be meant by different 'modes of thought' can only be a different linguistic thinking, i.e. one pertaining to the means of communication, not one that in any way constitutes meaning.

To distinguish his conception from the older Wundtian view, which, like Humboldt, subsumed a wide array of very different points under the general label *innere sprachform*, he refers to it as "figürliche innere Sprachform" / "figurative inner form," meant to refer only to the aspect of the notion discussed above and nothing else. It is here that the Wundtian notion of the dominating feature is at home (Marty 1908: 581, on the confusion of lexical motivation with semantics, see also Alinei 1997). Marty (1908) also elaborates on this notion, by addressing the differences and similarities found in the source concepts involved across languages (interestingly, Marty 1908: 177 already suggests the possibility that the relations of these to the target concept may be described by contiguity and similarity). As examples, he is adducing the cases of the different concrete sources for the meanings 'to think,' and, in a more detailed account, 'perhaps':

> Daneben aber besteht eine Großzahl von Fällen, wo die innere Form bei verschiedenen Sprachen anders und anders geartet ist und bloß der allgemeinsten Methode nach übereinstimmt. Ich erinnere beispielsweise an die von verschiedenen physischen Vorgängen hergenommenen Bilder, womit da und dort derselbe psychische Vorgang (vgl. „Denken" bald durch das Bild von einem Zusammenschütteln oder –bringen, bald von einem Wägen, bald von einem Teilen und Ausscheiden usw. usw.) umschrieben wird. ... daß, was wir durch „vielleicht" (= sehr leicht) ausdrücken, im Griechischen mit *ταχα* (hergenommen von *ταχύς*) oder *ἴσως* (hergenommen von den nach beiden Seiten gleichen

> Chancen), im Englischen mit perhaps (zusammenhängend mit hap, der Zufall), im Spanischen mit a caso (zusammenhängend mit casus), im Lateinischen mit forsitan (forssit an) wiedergegeben wird ... Diese Beispiele von inneren Sprachformen, die bei gleicher Bedeutung in verschiedenen Sprachen differieren und wofür sich die Beispiele ins Endlose vermehren ließen, sind besonders geeignet, die Verschiedenheit jener wechselnden Begleitvorstellungen von der überall identischen Bedeutung vor Augen zu führen, und umgekehrt der Fall, wo dieselbe innere Sprachform da und dort den Vermittler für verschiedene Bedeutungen bildet (1908: 141).

> Along with this, there is a big number of cases in which the inner form is diverse and differently natured and only concurs in the general method. I call to mind for instance the images, taken from different physical processes, by which here and there the same mental process is circumscribed (cf. 'thinking' by an image of shaking or bringing together, from weighing, or from separating or dividing etc. etc.). ... that that which we express by vielleicht (=very light) is rendered in Greek by *ταχα* (taken from *ταχύς*) or *ἴσως* (taken from equal chances on both sides), in English by *perhaps* (connected to *hap*), in Spanish by *a caso* (connected to *casus*[8]), in Latin by *forsitan* (*forssit an*)[9] ... These examples of inner forms that differ while the meaning is the same and for which examples could be multiplied ad infinitum are especially suited to bring home the diversity of the protean subsidiary conceptions of the semantics which is the same everywhere, and conversely the case in which the same inner form constitutes the facilitator for different meanings here and there.

However, in a short passage, Marty mentions that there even may be some truth to the original conception of *innere sprachform*, in that there indeed may be similarities in one language or language family, something like a conceptual fingerprint, when it comes to the selection of source concepts:

> Das Wahre an jener Rede von einer inneren Form in einer gewissen Sprache ist das, daß die dahin gehörigen Erscheinungen bei verschiedenen Ausdrucksmitteln derselben Sprache oder Sprachenfamilie unter sich vielfach Züge der Übereinstimmung, wo nicht etwas wie einen einheitlichen Stil zu zeigen pflegen (1908: 142).

> The truth in this talk about inner form in a certain language is that the phenomena pertaining to it tend to exhibit among themselves, considering the different means of expression in the same language or language family, traits of agreement, if not something like a uniform style, in many cases.

Unfortunately, Marty does not elaborate further on what precisely it is that he is alluding to; no examples are provided to illustrate his line of thought. For a more detailed overview of Marty's philosophy of language, see Funke (1924/1974).

---

[8] In fact the Spanish form is *caso*; *casus* as quoted by Marty is the Latin form to which it goes back. The *innere sprachform* here is that *casus*, in Latin, is connected to the verb *cadere* 'to fall,' i.e. the conceptualization of 'perhaps' here is one of (accidentally) falling somewhere at random.

[9] *Forsitan* is a contracted, lexicalized form of *forssit an*, which is in fact a phrase consisting of *fors* 'fate,' the 3rd singular conjunctive present tense form of the copula plus the subordinating conjunction *an*, and could be translated as something like "be the fate that..."

2.5. ONOMASIOLOGY

The transition from what may be called pre-modern linguistic reasoning to onomasiology is not as clear-cut as the division in sections made here might suggest. This may have something to do with the fact that onomasiology is essentially non-structuralist in nature and has therefore not experienced as dramatic changes in theory with the advent of structuralism (see Kramer 2000 for an outline of its development). Indeed, many onomasiological works retain the same underlying assumptions taken for granted by earlier writers like those discussed above. The basic task of onomasiology, in the words of Zauner (1902: 4), who also coined the term in the first place (although even earlier works such as Diez 1875 and Tappolet 1895 already assumed essentially the same perspective), is to take "den Begriff zum Ausgangspunkt" / "the concept as the starting point" and to determine "welche Bezeichnung, Benennung die Sprache für diesen Begriff habe" / "which denomination and designation the language may have for this concept," and in a second, more analytic step, "zu ergründen, warum die Sprache dieses oder jenes Wort zur Benennung dieses oder jenes Begriffes verwendet" / "to determine why the language utilizes this or that word to designate this or that concept." In their empirical work, Diez, Tappolet, and Zauner were, aside from outlining the new approach to the study of the lexicon, concerned with different semantic fields in Romance languages: Tappolet (1875) investigates kinship terms and Zauner (1902) body part-terms, also under diachronic aspects. Work in the onomasiological tradition in the first half of the 20th century, largely carried out by German-speaking scholars, has spawned a huge amount of literature investigating individual "bennenungsgründe"[10] (roughly, "naming rationale") for concepts in languages of a certain area or family (see Grzega 2009 for an extensive bibliography).

[10] In fact, this term is a fine example to illustrate how onomasiology inherited its notional apparatus from works at the dawn of modern linguistics. "Benennungsgründe" can at least be traced back to Bopp (1836: 136), who suggested that the purpose of etymology be

> daſs man, so weit es möglich ist, einem jeden Worte die Gesetzmäſsigkeit seiner Bildung nachweist, ihm gleichsam seinen Lebenslauf zur Seite stellt, sein Aussehen in früheren Perioden, d.h. in älteren stammverwandten Sprachen beschreibt, und durch die Zusammenstellung der sich wechselseitig aufklärenden Formen die echteste, ursprünglichste von allen ermittelt, und hierdurch häufig den Benennungsgrund eines Gegenstandes aufdeckt, und so einerseits die der Sprache innewohnende Philosophie, die Sinnigkeit ihrer Uranschauungen, und andererseits die Regelmäſsigkeit und Natürlichkeit ihrer physischen Einrichtung, so wie die einfachsten Elemente ihres Ganzen an das Licht zieht.
>
> that one, as far as possible, detects the regularity of each word's formation, and that one so to speak provides it with its vita, describes its appearance in earlier periods, i.e. in older genetically related languages, and determines, by compiling the forms that illuminate themselves mutually, the most original, pristine one of them all, and thereby one often lays bare the naming rationale of an object, and so sheds light on the one hand on the philosophy that indwells the language, the meaningfulness of its primordial conception, and on the other hand the regularity and naturalness of its physical composition, as well as the simplest elements of its whole.

Thus, it is clear that, for Bopp, the task of etymology is intricately connected to the question of how languages verbalize experience and its relevance for presumed differences in thought.

While onomasiological works are often characterized by particularism with respect to the meanings investigated and are restricted to individual case studies on a small scale, at least some writers at the same time outline broader research agendas. Indeed, Tappolet (1895: 2) suggests a new research branch he calls "vergleichende Lexikologie" ("comparative lexicology") which is based on onomasiological principles and seeks to elucidate naming motives in different languages for the same concepts. This is obviously a programmatic extension of the task already inherent in the approach of earlier writers, such as Pott.

2.6. STRUCTURALISM

Lexical motivation also had a role to play in structuralist thinking. It is relatively unknown when compared with the huge impact of the Saussurean doctrine of the arbitrariness of the linguistic sign that Saussure introduced the notion of the motivation of the linguistic sign into the scientific discourse at the same time. Saussure (1916/1967: 180-181) writes:

> Le principe fondamental de l'arbitraire du signe n'empêche pas de distinguer dans chaque langue ce qui est radicalment arbitraire, c'est-à-dire immotivé, de ce qui ne l'est que relativement. Une partie seulement des signes est absolutement arbitraire ; chez d'autres intervient un phénomène qui permet de reconnaître des degrés dans l'arbitraire sans le supprimer : *le signe peut être relativement motivé.* Ainsi *vingt* est immotivé, mais *dix-neuf* ne l'est pas au même degré, parce qu'il évoque les termes dont il se compose et d'autres qui lui sont associés ... Il en est de même pour *poirier*, qui rappelle le mot simple *poire* et dont le suffixe *-ier* fait penser à *cerisier, pommier*, etc. ...
>
> The fundamental principle of the arbitrary nature of the linguistic sign does not prevent us from distinguishing in any language between what is intrinsically arbitrary - that is, unmotivated - and what is only relatively arbitrary. Not all signs are absolutely arbitrary. In some cases, there are factors which allow us to recognise different degrees of arbitrariness, although never to discard the notion entirely. *The sign may be motivated to a certain extent*. The French word *vingt* ('twenty') is unmotivated, whereas *dix-neuf* ('nineteen') is not unmotivated to the same extent. For *dix-neuf* evokes the words of which it is composed, *dix* ('ten') and *neuf* ('nine') and those of the same numerical series ... The same is true of *poirier* ('pear-tree'), which evokes the simple form *poire* ('pear') and has a suffix *–ier* which recalls that of *cerisier* ('cherry-tree'), *pommier* ('apple-tree'), etc. (Harris 1983: 130)

According to Saussure, complex expression, such as *dix-neuf* 'nineteen' and *poir-ier* 'pear tree' are thus relatively motivated by virtue of the fact that their meaning is constituted in some unspecified manner. As the last above quoted passage suggests, this happens in what Saussure (1916/1967) calls "rapports associatifs," that is, by paradigmatic interconnections to signs that are either similarly formed, similar semantically, or both. Alinei (2001) raises the question as to who introduced the term *motivation* into linguistics. It is probable that Saussure did not invent the term, but rather was influenced by earlier writers. Koerner (1971: 165fn38) points to a similar usage of the term in Kruszewski (1890). The ultimate source of the term should, however, as it seems, be sought in philosophical writings widely read in intellectual circles around 1900. In Franz Brentano's (1956: 128) theory of mental judgements, one finds the statement: "Motiviert ist ein Urteil, wenn es unmittelbar von einem anderen psychischen Phänomen verursacht wird und wir diese Verursachung

wahrnehmen" / "a judgement is motivated if it is caused directly by another psychic phenomenon and if we perceive this causation." Given the psychological flavor of Saussure's account of relative motivation and the *rapports associatifs* and of Kruszewski's treatment of linguistic change, it seems likely that Brentano is the source from which Saussure, either directly or indirectly via Kruszewski, borrowed the notion of motivation and applied it to linguistic questions.

Even more important than the mere introduction of the idea of relative motivation, however, is that Saussure also outlined a rudimentary typology with respect to the degree to which languages utilize arbitrary and relatively motivated lexical items and suggests that this is an important property that may be used to establish typological groupings:

> Il n'existe pas de langue où rien ne soit motivé ; quant à en concevoir une où tout le serait, cela serait impossible par définition. Entre les deux limites extrêmes – minimum d'organisation et minimum d'arbitraire – ou trouve toutes les varietés possibles. Les divers idiomes renferment toujours des éléments des deux ordres – radicalment arbitraires et relativement motivés – mais dans des proportions très variables, et c'est là un caractère important, qui peut entrer en ligne de compte dans leur classement. En un certain sens – qu'il ne faut pas serrer de trop près, mais qui rend sensible une des formes de cette opposition –, on pourrait dire que les langues où l'immotivité atteint son maximum sont plus *lexicologiques*, et celles où il s'abbaise au minimum, plus *grammaticales* (Saussure 1916/1967: 183).

> There exists no language in which nothing at all is motivated. Even to conceive of such a language is an impossibility by definition. Between the two extremes –minimum of organisation and minimum of arbitrariness- all possible varieties are found. Languages always exhibit features of both kinds – intrinsically arbitrary and relatively motivated – but in very varying proportions. This is an important characteristic, which may have to be taken into account in classifying languages. In one sense –this must not be pressed too far, but it brings out one aspect of the contrast – a distinction could be drawn between lexicological languages, in which absence of motivation reaches a maximum, and *grammatical* languages, in which it falls to a minimum (Harris 1983: 131-132).

Saussure goes on to purport that German, when compared with English,[11] is closer to the lexicological pole on the continuum of lexicological versus grammatical languages. As an example of "l'ultra-lexicologique," that is, of an "ultra-lexicological" language, he cites Chinese, and as "spécimens de l'ultra-grammatical" / "exemplars of the ultra-grammatical" he cites Indo-European (apparently the Proto-Language is meant) and its early descendant Sanskrit. Interestingly, despite the purported largely ahistorical perspective of Saussurean structuralism, Saussure even suggests that languages may shift their position on the lexicological-grammatical continuum by diachronic change (see Urban 2008 for empirical diachronic data for Latin and Spanish):

[11] See Scheidegger (1981) for an evaluation of the alleged higher degree of motivated lexemes in the Saussurean sense in German when compared with French.

> Dans l'intérieur d'une même langue, tout le mouvement de l'évolution peut être marqué par un passage continuel du motivé à l'arbitraire et de l'arbitraire au motivé ; ce va-et-vient a souvent pour résultat de déplacer sensiblement les proportions de ces deux catégories de signes Saussure (1916/1867: 184).

> Within the same language, a whole evolutionary trend may be marked by constant movement from motivation to arbitrariness, and vice versa. The result of this to-and-fro is often a noticeable shift in the proportions of the two categories of sign. (Harris 1983: 132).

Saussure's typology *in statu nascendi* was taken up and elaborated on at various points by Stephen Ullmann, who is therefore sometimes credited to be "[o]ne of the founding fathers of lexical typology" (Koch and Marzo 2007: 260). Ullmann (1962) distinguishes three types of motivation: phonetic motivation (onomatopoeia and sound symbolism), morphological motivation (derivatives and compounds), and semantic motivation (metaphorical extensions, such as that from 'hood' to 'hood of car'). As for morphological motivation, Ullmann (1962: 91) remarks that "in many cases the connexion [sic!] between the two elements may be remote or obscure, as for instance in *butterfly*, *kingfisher*, or *lady-bird*, but it is none the less obvious that such words are morphologically motivated." Ullmann (1962: 91) concedes that particular words may be motivated in more than one way: "The plant name *blue-bell*, for example, has such mixed motivation: it is a transparent compound and at the same time a metaphor based on the bell-like shape of the flower." Ullmann (1962: 93) further makes clear that the locus of analysis is, or should be, judgements by speakers: "For a word to be so motivated, it must be felt to be a compound, a derivative, or a figurative expression. Once again it might be possible to devise a statistical method in order to determine, in marginal cases, how far people are aware, or can be made aware, of the motivation of such words." Ullmann thus broadens the Saussurean notion of relative motivation, which he identifies with his morphological motivation, by recognizing phonetic and semantic factors as distinct types of lexical motivation. Note, however, that Saussure's concept of relative motivation is not reducible to morphological factors alone in spite of their important role, but has, as seen, a distinct psychological component built into it. To this extent Ullmann also altered the original Saussurean conception in a significant way.

Ullmann is very interested in the typological aspect of the distinction between motivated and arbitrary words as formulated by Saussure and calls this "one of Saussure's most important discoveries" (1962: 105). He suggests differences between languages of the different types with respect to the degree to which loanwords are accepted into the language (1962: 112) and ponders the possibility of and reasons behind similar metaphorical processes in unrelated languages across the world (1966: 238). However, Ullmann, like Saussure, only cites random examples anecdotally to make a case for one language or another being relatively motivated or not, and restricts himself to suggestions for further work while at the same time noting potential difficulties with non-morphological motivation:

> [T]hough one may have some quite definite impressions about the frequency of onomatopoeia or metaphor in a given language, it would be difficult to formulate them with any

> degree of precision. With morphological motivation one is on firmer ground: it is the most clear-cut and least subjective of the three types, and certain broad tendencies stand out very clearly even though they may not be statistically formulable (Ullmann 1962: 105).

In a later publication, Ullmann goes one step further in specifying how one might go about testing the degree of motivated and arbitrary terms in a given language:

> It might be possible to devise some statistical test for these relative frequencies. Such a test might be based on samples from dictionaries, on a representative selection of texts, or on both. Such isolated numerical data as are already available seem to be very suggestive (1966: 223).

As is obvious, this is precisely the task of the present study!

Further structuralist work on lexical motivation includes Gauger (1971) and Rettig (1981) among others, which are not discussed here at length because they are not concerned with cross-linguistic questions or make theoretical contributions that are relevant to structuralist thinking in particular. An independent account of lexical motivation that takes Saussure's writings as its starting point is represented by Alinei (1996, 1997, 2001), among other publications. Alinei proposes that research into lexical motivation, for which he suggests the term iconymy, should form an autonomous subdiscipline of linguistics. Next to a short account of the role motivated terms have to play in language, Alinei (2001: 92-93) in particular offers a rendition of the by now well-known notion of choosing different aspects of the object for its linguistic designation:

> In the case of 'eyeglasses', for example, each iconym collapses and represents a sort of encyclopaedic [sic!] definition of 'glasses', which can be something like 'device consisting of a pair of crystal round lenses to improve human vision mounted in frames held on the bridge of the nose with sidepieces to grip the temples; originally beryl was used etc.' . Out of this definition, one can choose, arbitrarily, 'glass', 'crystal', 'beryl', 'hook', 'eye', 'lens' etc., as a condensed representative of the whole concept. Also French *lunettes* 'little moons' is based on one of the components of the encyclopaedic definition, namely 'round (lens)', but is metaphorical (associative) in nature, and not merely descriptive. Notice that the choice of iconyms is always arbitrary, as it is made from within a practically unlimited set of conceptual candidates and/or their metaphoric equivalents.

Next to the interesting claim that languages arbitrarily pick one or the other bit of the encyclopedic knowledge about the object to be designated, Alinei goes on to sharpen the theoretical distinctions between strategies in which the component chosen for the lexical designation "belongs to the structural paradigm of the designandum" and those belonging to a "different conceptual sphere," such as French *lunettes,* which are said to be "metaphoric in kind." A large-scale project initially headed by Alinei that is still ongoing is the *Atlas Linguarum Europae*, commencing with the publication of Alinei (1983). The aim of this project is to examine the denominations of an impressive variety of concepts across languages of Europe in the form of maps and accompanying datasets.

2.7. LINGUISTIC ANTHROPOLOGY

Lexical motivation, both in the form of polysemous and morphologically complex expressions, has a quite important role to play in the investigation of ethnobiological taxonomies, in particular in their historic developments, although other terms are traditionally used by researchers in this area to refer to the phenomena ("unitary labels" is typically used for simplex terms and "overtly marked" terms for morphologically complex expressions).[12] Berlin (1972), in a seminal paper, demonstrates that, when terms for erstwhile unlabeled higher ranks in the taxonomy, such as the life-form rank (corresponding to English *bird*) and the unique beginner rank (corresponding, roughly, to English *creature*), are developed in a language, these either arise in the form of semantic expansion (i.e. development of polysemy) of a lower-level term, or are (at least) initially expressed by morphologically complex terms.

Apart from the realm of ethnobiological classification, important research on regularities of lexical motivation, with an explicitly cross-linguistic orientation, was carried out by the anthropological linguists Cecil H. Brown and Stanley R. Witkowski, who published a series of studies in the early 1980s. Importantly, work by these scholars not only makes empirical observations, but usually also attempts to come up with an explanation for each pattern, which typically recurs to language-independent cultural factors.

Witkowski et al. (1981) argue that terms for 'tree' are relatively recent additions to the lexicon in many languages and that "thousands of years ago most languages lacked a 'tree' category" (Witkowski et al. 1981: 10). They hypothesize that the development of terms for this concept is triggered by increasing societal complexity. Linguistically, they suggest that 'tree' terms arise as an additionally encoded meaning of terms for 'wood' with "low salience" and present some cases of morphologically complex terms for 'tree' based on 'wood' which they interpret to the effect that this "may constitute an incipient phase in the separation of 'wood' and 'tree' referents" (Witkowski et al. 1981: 9). Similar points are made for languages which conflate lexically the meanings 'eye' and 'face' and 'seed' and 'fruit' respectively in Brown and Witkowski (1983).

Witkowski and Brown (1985) examine the areal distribution of languages without lexical differentiation between 'hand' and 'arm' and 'foot' and 'leg' respectively, and establish that such languages dominate in regions close to the equator. They (1985: 207) speculate that "an important influence on the occurrence of limb polysemy in languages is the existence of extensive wearing apparel in societies, especially tailored clothing covering the limbs and other limb gear" (see also Brown 2005b). Brown (2005a) investigates this situation with respect to the referents 'finger' and 'hand' and suggests that languages without any lexical differentiation for these referents are typically hunter-gatherers societies because they "differ from agrarians in the extent to which they make use of finger adornment," which is taken as an explanation of the observed patterns.

Further, Brown and Witkowski (1981) examine metaphorical denomination strategies (which they refer to as "figurative") for certain parts of the body, such as 'finger' and 'toe,' 'pupil of the eye,' 'muscle,' and 'testicle.' More than other studies, this one focuses on

[12] Berlin (1992), however, discusses the relation of ethnobiological nomenclature with the traditional distinction between arbitrary and motivated.

the striking cross-linguistic similarity of these metaphor-driven strategies (observed are "child of hand/foot" for 'finger' and 'toe,' "person of the eye" for 'pupil,' "mouse or other small animal (of the arm)" for 'muscle' and "egg" for 'testicle') and possible approaches to their explanation. In addition to lexical constraints, Brown and Witkowski (1981: 606) note that "[t]he limited number of things in the physical world that resemble or are in some way regularly associated with body parts significantly constrains the types of figurative equations that can achieve currency in a language," and, in discussing why metaphorical denominations as opposed to non-metaphor-driven complex terms are preferred, claim that the former are "more interesting, fetching, and dramatic" (1981: 607). Brown (1999) examines lexical acculturation, i.e. the process of naming for novel entities encountered in the course of contact with European culture in Native American languages, and likewise notes striking similarities across languages with respect to the linguistic treatment of these entities. Brown (1999) appeals, with reference to work by Chomsky, to a "detailed wiring" approach in cognition as the main explanatory theory, which is, however, not described in great detail. Finally, Brown (1983) examines origins in the words for the cardinal directions across languages, and Witkowski and Brown (1983) demonstrate how "marking reversal," i.e. the situation in which one referent that was originally designated by a simplex term comes to be expressed by a complex term under the influence of introduction of a new referent that needs to be named, can bring about lexical semantic change. For a concise summary of research in this paradigm, see also Brown (2001).

2.8. MAVERICKS

2.8.1. ZEHETMAYR

Zehetmayr published an "analogically-comparative" dictionary of Indo-European (with an emphasis on Latin, Greek, Sanscrit and Germanic languages) in 1879. Under the label *analogy* Zehetmayr subsumes two different notions: formal analogy on the one hand, by which he understands forms shared by related languages due to common descent as well as formal similarities (e.g., when terms in two languages are formally derived by means of an affix or the like), and "analogy of ideas" on the other. Zehetmayr assigns greater importance and value to the latter, both because of their practical value for etymological research as well as because of their value for something like a "natural philosophy" enshrined in language, an idea also encountered in the writings of Pott; this position is expressed in the following quote from Zehetmayr (1879: iii-iv):

> Ungleich wichtiger, als die sprachlich *formale* Analogie, ist die *Ideen*-Analogie, welche es mit dem Grundbegriff zu thun hat, den selbst etymologisch nicht verwandte Wörter aus verschiedenen Sprachen in der Bezeichnung eines Gegenstandes gemeinsam theilen. Die Ideen-Analogie ist die exclusiv philosophische Seite der Linguistik, indem sie auf die primitive Identität im Denken, d.h. in der Vernunft, und noch weit mehr, als die Stammverwandtschaft im äussern Wortlaut, womit sich die Etymologie beschäftigt, auf die Einheit unseres Geschlechtes schliessen lässt. Ja, noch mehr: die philosophische Definition nicht weniger Begriffe gewinnt in der Ideen-Analogie einen soliden, sichern, weil positiven Boden, statt in oft schwankenden Subtilitäten sich zu bewegen, so dass sie für Viele nicht

> zu einem befriedigenden Abschluss gelangt. ... wo die nach Ort und oft auch nach Zeit von einander entferntesten Völker in der Ideenfassung durch das Wort harmoniren, herrscht so zu sagen Unfehlbarkeit ... Die Ideen-Analogie der Sprache bietet Ueberraschendes in solcher Fülle, dass die Linguistik, die ja selbst wieder für ihre Forschungen bezüglich mancher etymologisch noch nicht feststehender Wörter durch jene Gedanken-Aehnlichkeit auf den oft einzig sichern Standpunkt versetzt werden kann, von wo as sich in Ruhe weiter operiren lässt, unmöglich von Versuchen Umgang nehmen kann, die, wie die vorliegende Schrift, der Ideen-Analogie vorzugsweise ihre Aufmerksamkeit zugewendet hat.
>
> Much more important than formal linguistic analogy is the analogy of ideas, which has to do with the basic concept, which even etymologically unrelated words from different languages have in common in the denomination of an object. The analogy of ideas is the exclusively philosophical side of linguistics, by virtue of which it is possible to deduce primitive identity in thinking, i.e. in reason, and far beyond genetic affinity in the outer shape of words with which etymology deals, the unity of our race. Even further: the philosophical definition of numerous notions attains solid, secure, since positive ground, instead of moving in often unsteady subtle consideration, so as to not reaching a satisfactory completion in the eyes of many. ... where the most remote peoples in terms of location and often also in time harmonize in the conception of ideas by means of the word, there is infallibility, so to speak ... The analogy of ideas of language offers surprising facts in such abundance that linguistics, which itself in turn for its research regarding some words not yet etymologically accounted for can only be put on safe grounds by this similarity in thought, can impossibly dodge attempts which, like the present volume, called attention to the analogy of ideas.

Despite not being on the methodological level of modern Indo-European studies due to its age, Zehetmayr's work offers an impressive amount of data that demonstrates that even the study of one language family can yield interesting semantic parallels that do not appear to be due to common descent in all cases.

2.8.2. SCHRÖPFER

Schröpfer (1979) is the first volume of a monumental, but unfinished, project of a comparative dictionary similar in many ways to Zehetmayr's work as well as to Buck's (1949) Dictionary of Selected Synonyms in the Principal Indo-European languages. The first published fascicle is based on data for 28 European languages, deliberately including some that were not featured in Buck's work (Schröpfer 1982: 159), but the dictionary was meant to include in its final form data from as many as 87 languages, including non-Indo-European ones (Schröpfer 1979-1994: xxiv). Its final version was planned to include 3,000 headwords, roughly double the number of Buck (1949). The purpose of such a *vergleichende onomasiologie*, or "comparative onomasiology," which is the programmatic motto chosen by Schröpfer for the dictionary, was to produce a repository of semantic associations and changes not only as a valuable research for etymological research, but also to demonstrate "daß die Benennungsvorgänge, die mit der menschlichen Wahrnehmung-Begriffsildung innig zusammenhängen, sich in allen Bereichen der Sprache und in allen Zeiten und Räumen ihrer Entwicklung und ihres Gebrauchs wiederholen und Analogieschlüsse zulassen" / "that the processes of denomination, with which the constitution of human perception and concept formation are intimately connected repeat themselves in all areas of languages and in all times and places of its development and usage and allow for conclu-

sions by analogy" (Schröpfer 1974b: 4) as a purpose in itself. In connection with this purpose, it is interesting to note that Schröpfer (e.g. 1979: xlv) uses the Humboldtian term of *innere sprachform*, and that not simply to establish a historical connection with his predecessor, but rather as a matter of course and as a technical term within his framework in connection with the "Frage der Übereinstimmung der Benennungsweise ..., d.h. ... die Frage zugleich der inneren Sprachform" / "question of agreement of the mode of denomination, i.e. simultaneously the question of the inner form of language." The dictionary, unlike Buck (1949), was also meant to feature an analytical framework and not merely list individual denominations. This framework is otherwise heavily influenced by the paradigm of *sprachinhaltsforschung* (e.g. Weisgerber 1971), and features, according to Schröpfer (1982: 163), the four parameters of "Nennwert/Nennfunktion," i.e. the lexical semantics of individual lexemes themselves, "Deutwerte/Deutefunktion," which is meant to represent den "ursprüngliche[n] Versuch, den Gegenstand nach einem später oft verlorenen oder vergessenen Merkmal zu benennen" / "the primordial attempt to name the object for a feature often lost or forgotten at a later point," i.e. the *innere sprachform* inherent in a transparent or etymologizable word, so-called "Nebenvorstellungen," a concept which is again directly taken over from Humboldt himself, as well as a lexeme's "Gefühlswert." As will be noted, the terminology employed by Schröpfer has an archaic feel to it, and it was probably very conservative already even at the point of time the dictionary was conceived of. Nevertheless, Schröpfer's purpose is quite clear and places him in line in a long, albeit somewhat discontinuous tradition. For a condensed outline of the project in English, see Schröpfer (1974a).[13]

[13] Schröpfer points in various papers to an awareness of differences in lexical motivation and the demand to systematize them on the side of philosophers, referring, among others, to Nietzsche, but without offering more specific references. With respect to Nietzsche, Schröpfer might be referring to the following passage from Nietzsche (2005/1873):

> Wir teilen die Dinge nach Geschlechtern ein, wir bezeichnen den Baum als männlich, die Pflanze als weiblich: welche willkürlichen Übertragungen! Wie weit hinausgeflogen über den Kanon der Gewißheit! Wir reden von einer 'Schlange': die Bezeichnung trifft nichts als das Sichwinden, könnte also auch dem Wurme zukommen. Welche willkürlichen Abgrenzungen, welche einseitigen Bevorzugungen bald der, bald jener Eigenschaft eines Dinges! Die verschiedenen Sprachen, nebeneinander gestellt, zeigen, daß es bei den Worten nie auf die Wahrheit, nie auf einen adäquaten Ausdruck ankommt: denn sonst gäbe es nicht so viele Sprachen. Das 'Ding an sich' (das würde eben die reine folgenlose Wahrheit sein) ist auch dem Sprachbildner ganz unfaßlich und ganz und gar nicht erstrebenswert. Er bezeichnet nur die Relationen der Dinge zu den Menschen und nimmt zu deren Ausdrucke die kühnsten Metaphern zu Hilfe. Ein Nervenreiz, zuerst übertragen in ein Bild! Erste Metapher. Das Bild wieder nachgeformt in einem Laut! Zweite Metapher. Und jedesmal vollständiges Überspringen der Sphäre, mitten hinein in eine ganz andre und neue.
>
> We divide things according to their genders; we designate the tree as masculine, the plant as feminine: what arbitrary metaphors! How far flown beyond the canon of certainty! We speak of a 'serpent'; the designation fits nothing but the sinuosity, and could therefore also appertain to the worm. What arbitrary demarcations! what one sided preferences given sometimes to this, sometimes to that quality of a thing! The different languages placed side by side show that with words, truth or adequate expression matters little: for otherwise there would be not so many languages. The 'thing-in-itself' (it is just this which would be the pure ineffective truth) is also quite incomprehensible to the creator of language and not worth making any great endeavor to obtain. He designates only the relations of things to men,

2.8.3. EILERS

Another author with a very clear conception of the theoretical-philosophical and practical-philological value of cross-linguistic research on lexical motivation is the iranianist Wilhelm Eilers. Given the striking commonalities between Eilers's and Schröpfer's approach that will become clear below, it is worth stressing that they seem to have been entirely unaware of each other's work in spite of the fact that they were publishing their results at roughly the same time. Perhaps surprisingly, Eilers calls his research paradigm, which is already foreshadowed in Eilers (1967), *vergleichende semasiologie* (Eilers 1973a,b), which at first glance suggests a very different conception when compared with Schröpfer's *vergleichende onomasiologie*, given that semasiology and onomasiology are traditionally conceived of as converse viewpoints that can be taken in analyzing linguistic signs (see e.g. Koch 2001). However, the different names Schröpfer and Eilers choose only very superficially conceal the fact that they essentially tackle the very same question, with Schröpfer putting emphasis on the semantics-based (onomasiological) approach to linguistic comparison and Eilers on the commonalities in conceptualization by virtue of the linguistic structure of the respective terms that is revealed by such a comparison. Like Schröpfer, Eilers emphasizes the remarkable finding of commonalities in the semantic associations found in very different languages:

> Es ist die Tatsache, daß zwei oder mehrere oft völlig verschiedene Sprachen sich zum Ausdruck eines und desselben Gedankens (Satz oder Wort) einer ganz gleichen oder mindestens doch ähnlichen Vorstellungsweise bedienen. Entlehnung darf dabei nicht im Spiele sein. Handelt es sich doch geradezu um ihren Gegensatz … : selbstständige Entstehung gleicher kultureller Phänomene an voneinander unabhängigen Orten, aber unter gleichen inneren und äußeren Vorraussetzungen (1973b: 11).

> It is the fact that two or even more completely different languages utilize a completely identical or at the least similar conceptualization to express one and the same thought (sentence or word). Borrowing must not be involved here. In fact it is a matter of the opposite: autonomous development of similar cultural phenomena in places that are independent of one another, but under similar inner and outer conditions.

> Diese höchst bemerkenswerten Ähnlichkeiten der Ausdrucksweise der voneinander räumlich wie zeitlich, besonders aber auch morphologisch und genealogisch entferntesten Sprachen drängt zu der Annahme hin, daß der Sprache als solcher überall, wo Menschen sprechen, gemeinsame Grundzüge innewohnen, daß die Sprachen im philologischen Sinn eine große Einheit bilden, die die Einheitlichkeit des Menschengeistes von den primitivsten Stammesverhältnissen in Afrika und Australien bis zur letzten Hochkultur der Völker Asiens und Europas unwiderleglich dartun (1973a: 10).

> These highly remarkable similarities in the modes of expression of spatially and temporally, and especially morphologically and genealogically most separated languages impels

---

and for their expressions he calls to his help the most daring metaphors. A nerve-stimulus, first transformed into a percept! First metaphor! The percept again copied into a sound! Second metaphor! And each time he leaps completely out of one sphere right into the midst of an entirely different one (Mügge 1964: 177-178).

> one to the assumption that common main features inhere in language as such everywhere people speak, and that languages form a large unity in the philological sense, and they substantiate irrefutably the unity of the human mind from the most primitive tribal conditions in Africa and Australia up to the last advanced culture of the peoples of Asia and Europe.

Also like Schröpfer, Eilers emphasized the methodological value of a *vergleichende semasiologie* for etymological research:

> Mit Hilfe der semasiologischen Methode lassen sich etymologische Ableitungen aufs neue überprüfen, möglich und wahrscheinlich machen oder eben ganz zurückweisen ... Kehrt die gleiche oder die ähnliche semantische Entwicklung in den verschiedensten Sprachen öfter wieder, so läßt sich auf eine der Menschheit oder Teilen von ihr gemeinsame Vorstellungsweise schließen, und wir dürfen sie auch anderswo erwarten (1973a: 22).

> By means of the semasiological method, etymological derivations can be reassessed anew, made possible and likely or be refuted entirely ... If the same or a similar semantic development recurs in very different languages repeatedly, then one can deduce a common mode of conception of mankind or parts of it, and we are entitled to expect it elsewhere as well.

Eilers also produced valuable empirical results mainly based on comparing Indo-European languages with Semitic and languages of the Middle East, in spite of a questionable assumption of the primacy of the abstract over the concrete in language on his behalf (Eilers 1973b: 6) in the light of modern research on diachronic semantics and grammaticalization, which has amassed data suggesting that precisely the opposite direction is preferred. Unlike Schröpfer, Eilers also frequently adduces evidence arrived at by means of etymological reconstruction. Among the commonalities across language families, either in the form of synchronically transparent lexical motivation or etymological lexical connections, Eilers (1973a) mentions the following: 'lungs' – 'light,'[14] 'eye' – 'to see,' 'ear'– 'to hear,' 'liver' – 'fat, heavy,' 'neck' – 'to turn,' 'ring' – 'finger,' 'mirror'– 'to see, to look,' 'tree'– 'to erect, to build,'[15] 'garden' – 'to enclose,' 'soul' – 'breath, puff,' 'wick' – 'to twine,' 'area, region' – 'circle,' 'thing' – 'property' – 'wish, desire,' 'thing' – 'word,' 'river, sea' – 'to gleam,' 'sun/star' – 'to burn,' 'tree' – 'firewood' – 'to burn,' 'elm' – 'fly/insect-tree.' It is not possible to either confirm or reject all of those suggestions simply because the sets of investigated meanings are only partially overlapping, but some of Eilers's suggestion, such as the connection between the 'lungs' and 'light' can certainly be confirmed as robust (see Appendix E, 122), while for others there is little evidence; for instance, there are no instances of languages in the sample to be described in chapter 3 where the word for 'river' or 'sea' is synchronically clearly derived from a verb meaning 'to gleam.'

[14] In the sense of 'not heavy.'

[15] See Turner et al. (1998: 387) for an approximate Salishan parallel.

2.8.4. SAPIR

Edward Sapir was well aware of differences between languages in the amount of motivated terms they employ, and how such differences might be exploited for linguistic analysis. Sapir (1916/1949: 435) draws attention to the fact that motivated terms are likely to be more recent coinages than unanalyzable words which, "through the destructive agency of gradual phonetic change," will tend to lose their motivated character, should they ever have had one. Sapir employs this observation to determine "the relative ages of cultural concepts" among communities of North America as one technique of the overall goal Sapir sets for himself in the article, namely to put forward methods to uncover cultural relations between North American communities and to assess their time depth.

However, lexical motivation also seems to have played a role in Sapir's thinking about the relation of language to culture and vice versa. The following quote from a recently reconstructed lecture series is illuminating in this context:

> In two languages one may find the form (sound) and the function (meaning) of elements to be the same but the patterns totally different. It is the internal economy –the configurational analysis- that is completely different in all languages. Suppose, for example, that in language A, the form *wala* means 'house' and in language B there is also a form *wala* meaning 'house'. Yet although the two forms are linguistically and culturally the same they can still be significantly different. Why? Because there may still be a difference in the morphology or configuration of the languages. In language A, *wala* consists of *wa* + *la*. *wa* means 'to dwell' and *la* means 'that which is used'. In language, B, however, *wala* is composed from *w-* + *ala* (where *ala* = 'house' and *w-* is a prefix marking neuter gender). Thus the two forms are functionally different in the two languages ... Do meanings, as located in the world and its physical characteristics, explain the linguistic configurations in which people talk about them? Although the exigencies of adjustment to the world are fairly uniform –hunger and the search for food, etc. – the languages about these necessities are very different. Meaning or reference are articulated by speech – we don't know the world before we have speech. If we don't have symbols, we don't have meanings (Sapir 2002: 107, indication of editorial additions removed).

The relation of motivated (descriptive) terms with the external environment is also briefly discussed in Sapir (1912). In any case, it is worth noting that lexical motivation as at least an important part of *innere sprachform*, if not largely identical with it, plays a role in the writings of two of the most prominent authors, Humboldt and Sapir, associated with the coming into being of the idea of linguistic relativity. It seems safe to say that differences in conceptualization by means of lexical motivation thus played a hitherto undervalued role in the shaping of conceptions of linguistic relativity.

2.8.5. DESCRIPTIVITY (SEILER 1975)

Seiler (1975) formalizes the notion of "descriptive" words, which has been around as a term used rather informally to refer to a certain kind of analyzable terms at least since Sapir (1912). He observes that

> [e]s läßt sich nun in vielen Sprachen beobachten, daß [Namen] für Gegenstände des Denkens und der Welt, einschließlich Personennamen, von [Prädikaten] in ihrer Oberflächenerscheinung, also Verben hergeleitet sind. In dem Maße wie diese Herleitung

> eine direkte ist, sind die Benennungen 'deskriptiv'; je indirekter bzw. undurchsichtiger die Herleitung, desto weniger 'deskriptiv' bzw. mehr 'etikettierend' ist die Benennung (1975: 15).
>
> In many languages it can be observed that names for objects of thought and the world, including personal names, are derived from predicates in their surface structure, i.e. from verbs. To the extent that this derivation is a direct one, the denominations are 'descriptive;' the more indirect or opaque the derivation, the less 'descriptive' or more 'labeling' the denomination is.

The connection to the Saussurean dichotomy of arbitrariness and motivation is, as noted by Seiler (1975: 38) himself, obvious, although Seiler allows a rather fluent continuum between more or less descriptive terms. However, Seiler's notion of descripivity is not equal to sheer morphological complexity, but has two factors built into its definition: (i) the requirement that descriptive terms be derived from underlying predications and hence morphologically from verbs, and (ii) the possibility of a compositional interpretation as a prerequisite for a term to be called descriptive in this sense, while at the same time showing restrictions in its denotational range (Seiler 1975: 45-46). Interestingly, and apparently independently of Saussure, Seiler (1975: 38-39) also alludes to a purported higher degree of descriptivity in his sense in older stages of Indo-European and proposes that highly descriptive languages (as some of the early Indo-European languages) often lack a copula, while those which are less fond of descriptive denominations are more likely to feature one (as most of the modern Indo-European languages).[16] Seiler also briefly discusses the consequences of a high degree of descriptivity in the nominal lexicon to the overall lexicological organization of a language:

> Wenn, wie im Cahuilla, ein offenbar beträchtlicher Teil des Gesamtwortschatzes sich 'deskriptiv' aus primitiven Bestandteilen aufbaut, so kann vermutet werden, daß die Anzahl dieser primitiven Terme, aufs Gesamtvokabular gesehen, geringer ist als bei einer Sprache, die das 'deskriptive' Prinzip weniger stark bevorzugt (1975: 50).
>
> If, like in Cahuilla, an apparently considerable portion of the total vocabulary is made up 'descriptively' from primitive parts, then one can conjecture that the number of these primitive terms, with respect to the total vocabulary, is smaller than in a language that favors the 'descriptive' principle less strongly.

There are a number of subsequent studies that work with the notion of descriptivity as outlined by Seiler: Ultan (1975, 1976) proposes a number of metrics to determine the degree of descriptivity in different languages in the domain of body-part terms, and Walter (1976) examines deverbal derivation in German from this point of view; for a redefinition of the notion see Urban (2008).

[16] Nichols (2010) argues for a heavily verb-based lexicon in Proto-Indo-European and hence a "descriptive" nominal lexicon.

2.9 COGNITIVE APPROACHES

The term motivation has come to be used to refer to a number of sometimes only loosely related phenomena[17] (see Radden and Panther 2004a for an overview of different meanings of the term motivation in recent work). Motivation in a broader sense from the perspective of Cognitive Linguistics is the topic of Radden and Panther (2004b), thus considerably broadening the application of the term motivation beyond lexical motivation proper, which was at the core of Saussure's usage of the term. Radden and Panther (2004a: 1) assert that the Saussurean conception of relative motivation "is in the spirit of cognitive linguistics," and provide, taking Saussure as their starting point, an example of a Cognitive Linguistic analysis of compounding, using words for 'screwdriver' from eleven European languages as examples. They point out (2004a: 5) that "[c]ompounds are especially interesting complex expressions in that they are conventional names that highlight conceptual parts of a more complex conceptualization" and observe, applying Lakoff's (1987) Idealized Cognitive Model approach, that the conceptual means "that are chosen for naming purposes may vary from language to language." Thus, they find that languages select only a few of the possible elements in the proposed Idealized Cognitive Model for 'screwdriver': many select the concept 'screw,' such as English *screwdriver* and Italian *cacciavite*, (which is analyzed as a metonymic relation by Radden and Panther 2004a), others select coordinate concepts in the common domain of 'tools' in addition, such as Swedish *skruvmejsel*, literally "screw-chisel" (Panther and Radden 2004a point out the similarity in shape between chisels and screwdrivers), while Portuguese has a name for 'screwdriver' analyzed by Radden and Panther (2004a: 7) as involving metaphor: *chave de fenda* is literally "key of cut." Note that Radden and Panther's analysis, while cast in the modern terms of Cognitive Linguistics and applying its analytic apparatus, is at its essence a modern rendering of the old pretheoretical observation that is by now so familiar: selection of salient features for denomination with cross-linguistic variation as to these features, which is combined, as e.g. in Marty (1908), with an analysis in terms of contiguity-driven metonymy and similarity-driven metaphor, but here as part of a more general theory of Cognitive Linguistics.

2.10. CHAPTER SUMMARY

Leading up to the final discussion of Radden and Panther (2004a), the idea of cross-linguistic differences with regard to aspects (either quantitative or qualitative) of lexical motivation was traced through five centuries, and a striking similarity across authors with respect to their thoughts about the topic emerged, although they seem to have been largely unaware of each other.

While the present study again will use a novel approach to lexical motivation that will be outlined in the following chapter, one of its central concerns is precisely to put these casual observations on a more systematic and empirically sound cross-linguistic

[17] Recently, the topic of cross-linguistic aspects of lexical motivation has also received renewed attention, with a focus on diachronic patterns of semantic association (Zalizniak 2008, Hénault-Sakhno and Sakhno 2005, who stick to a redefined version of the Humboldtian notion of *innere sprachform* and build on diachronic work discussed in Sakhno 1999). See also Stéphane (1997) for a different account of the phenomenon.

basis. In this sense, this study joins its predecessors. The starting point to this endeavor will be the most outstanding contribution to a systematic study of lexical motivation from a cross-linguistic point of view in recent times, namely the approach developed by Koch and colleagues (most prominently Koch 2001 and Koch and Marzo 2007). Building on the concept of motivation as developed by Saussure and modified by Ullmann, Koch and his colleagues have developed a complete framework for the analysis of lexical motivation both in synchrony and diachrony, introducing both a more elaborate version of the formal aspects of lexical motivation and incorporating recent ideas from Cognitive Linguistics into the analysis of the semantic aspects of lexical motivation. While the treatment of the history of thoughts about cross-linguistic aspects of lexical motivation would have been incomplete without briefly mentioning the approach of Koch and colleagues, it will be discussed at length in the following chapter, which introduces the framework of the present study and sets out the basic classificatory grid used.

Chapter 3
The Analytic Framework of this Study

3.1. INTRODUCTION

In chapter 1, the term lexical motivation was introduced as a cover term to characterize morphologically complex and polysemous lexical items. Further, the potential results that can be expected from a systematic cross-linguistic comparison of the lexicon of the world's languages were sketched. Chapter 2 has demonstrated the role cross-linguistic aspects of lexical motivation have played in many intellectual currents in the past.

This chapter defines lexical motivation in more detail and introduces the particular fashion in which the concept is approached from a cross-linguistic perspective. In addition, the peculiar methodological challenges that accompany such a task are discussed. § 3.2. introduces how the notion of lexical motivation is operationalized so as to make its particular aspects cross-linguistically comparable; more specifically, the approach based on a list of meanings is described. § 3.3. describes the sampling procedure of the present study, § 3.4. makes explicit some general assumptions about the nature of the lexicon adopted, and § 3.5. defines the notion of lexical motivation, which has already been informally introduced in chapter 1. § 3.6. forms the heart of this chapter. It describes in detail the typological grid of form-meaning pairings that will serve as the background for subsequent typological comparison, and it describes the heuristic procedures used to classify a particular lexical item in a particular language in this grid. In addition, some serious theoretical problems of semantic analysis that are related to this endeavor are discussed, as well as the ways they are dealt with in the present study. § 3.7. moves to the more practical side of things in that it lists a number of further analytical decisions that had to be made; these relate mainly to the particular way the data were generated for the present study.

3.2. ESTABLISHING COMPARANDA

3.2.1. The meaning list

Chapter 2 demonstrated that differences in the quantity of motivated terms were noted early on, and it was pointed out that these differences played a certain role in structuralist thinking. Ullmann (1962) commented on difficulties in quantifying intuitions about these differences across languages, but later on (Ullmann 1966: 223) suggested that "it might be

possible to devise some statistical test for these relative frequencies. Such a test might be based on samples from dictionaries, on a representative selection of texts, or on both." Dictionary samples would indeed be a possible starting point to get a first impression of the behavior of different languages, but it is hard to arrive at a principled way of carrying out this kind of sampling and to systematize the results. One option would be to select every x^{th} page of a dictionary and to analyze the vocabulary items found there. This was done by Nettle (1995) and Pawley (2006) for different purposes. This approach, however, can yield an at best impressionistic overview, as Pawley himself notes. Analysis of texts, as suggested by Ullmann, would be very interesting as well. However, the texts to be analyzed would require to be glossed throughout to assess the degree of lexical motivation by morphological complexity, and such materials are not readily available for many languages, in spite of the growing number of corpora. Even more importantly, analysis of lexical items that occur in textual contexts make it intricately hard to quantify the degree of lexical motivation by lexically entrenched polysemy, because the embeddedness of lexical items in context typically selects only one of the possible readings that a lexical item might assume when it occurs in a different context (compare Cruse's 1986 discussion of contextual sense modulation). Texts were analyzed, however, for a much smaller set of languages (English, Latin, Spanish, Tuscarora) for which high quality textual data was available in addition to a fixed wordlist in Urban (2008), and it was found that there was a strong correlation between the two types of data for any of these languages.

Given that neither dictionary samples nor text analyses seem feasible, this study aims at quantifying the degree of lexical motivation in different languages utilizing a relatively small fixed wordlist of 160 meanings. The meanings are provided in English, but it should be stressed that the list is not to be thought of as a list of English words, but as a list of meanings. Most of the meanings under investigation are also found in the World Loanword Database (Haspelmath and Tadmor 2009c), and their semantics are elaborated in this publication by providing short sentences for a typical context in which they might naturally occur or by providing a definition.[1] These clues to the semantics of the items on the word list were adopted for the present study, and they served as guidelines for deciding whether a given lexical item in a particular language was to be considered semantically equivalent (see Haspelmath and Tadmor 2009b: 8). In some cases, it was necessary to narrow down intended meanings beyond Haspelmath and Tadmor. A fine example is the meaning 'flower,' for which no meaning description or typical context is given in the World Loanword Database. However, English *flower* is ambiguous and can either refer to the reproductive structures of plants (this reading is near-synonymous with *bloom* and *blossom*) or to a small plant with a decorative flower in the first sense (the meanings are treated distinctly for instance in German, which has *Blüte* and *Blume*). Presently, the first ('Blüte') sense of *flower* is targeted, since it is this sense for which it can be assumed that there is a conventionalized term in many languages. Furthermore, Buck (1949) was used as a guideline when it came to accepting or rejecting near-synonymous meanings as semantic proxies for the target meaning.

[1] This list is in turn based on Buck (1949).

The data acquisition process allowed for the relationship of meanings and lexical items to be many-to-many: rather than trying to pick the "best" equivalent for a given meaning in a target language, all lexical items listed under the respective headword in the consulted sources were taken into consideration and copied to the database for this study. Conversely, a single lexical item may correspond to two or even more meanings on the wordlist.

The 160 meanings belong to four semantic domains: topological and nature-related terms, artifacts, body-part terms, and finally, terms for phases of the day and a small number of miscellanea. In the following, each semantic domain, the individual meanings subsumed under it, and the reasons for its investigation with respect to lexical motivation are briefly described.

3.2.1.1. Topological and nature-related terms

1. animal
2. ashes
3. bark
4. bay
5. beak
6. bird
7. bloom (blossom, flower)
8. branch
9. bud
10. cave
11. clearing
12. cloud
13. coal
14. coast
15. dew
16. dust
17. eclipse
18. egg
19. embers
20. estuary
21. feather
22. flame
23. flood
24. foam
25. fog/mist
26. forest
27. gold
28. grass
29. headland
30. honey
31. horizon
37. Milky Way
38. moon
39. mountain
40. mushroom (fungus)
41. nest
42. plant
43. puddle
44. rain
45. rainbow
46. resin
47. river/stream
48. river bed
49. root
50. seed
51. shadow
52. sky
53. smoke
54. soil
55. spark
56. spring/well
57. star
58. steam
59. straw
60. sun
61. swamp
62. tail
63. thorn
64. thunder
65. tree
66. valley
67. volcano

32. horn
33. lagoon
34. lake
35. lightning
36. meteoroid (shooting/shining star)
68. waterfall
69. wave
70. wax
71. whirlpool

As Mark and Turk (2003a) point out, "[t]he landscape, a place to stand, places to live and find resources, is ... absolutely essential to human existence," and this putative basicness of landscape and natural phenomena to human experience is what makes them interesting semantic fields for the present study (see also Levinson 2008: 257). However, it would be a premature conclusion to believe that this basicness automatically entails that languages respond to the stimuli provided by the environment in uniform ways. Smith and Mark (2001) report relatively uniform responses in tests in which college students in the United States were asked to name "kinds of geographic features." The most frequent answers were items such as *mountain, river, lake, ocean, valley*, and so forth. As "preliminary data from parallel experiments carried out in Finland, Croatia, and the United Kingdom produced very similar trends" (2001: 610), Smith and Mark suggested that the domain of geographical features form a "coherent knowledge domain" with fairly uniform structure cross-linguistically. Mark and Turk (2003b: 39), however, in a study of landscape categories in Yindjibarndi, a Pama-Nyungan language of Australia, later found that "at the basic level of category terms, the Yindjibarndi landscape vocabulary is completely different from the terms covering the equivalent domain in English." The authors therefore suggested the initiation of a new research field of *ethnophysiography*, devoted to describing differences in the conceptual organization of landscape categories in different communities. Subsequently published data on a variety of languages in Burenhult (2008a), in particular Levinson (2008), have reinforced the need for the recognition of differences between individual languages. For instance, in Yélî Dnye, *mbu*, which can be used to refer to a 'mountain,' "is unspecified as to size, being applicable to features of varying magnitude (mountains, hills and even crab mounds on the beach), and only encodes that the feature has a conical shape" (Burenhult and Levinson 2008: 141). Perhaps more importantly, Levinson (2008) argues that categorization of the landscape domain need not be driven by perceptual salience of the environment's features. Instead, categorization is said to be often governed by human affordance (a term coined by Gibson 1977, 1979 to describe the latent possibilities for action the environment offers) and culture-specific models. On the other hand, in Jahai, as described by Burenhult (2008b: 185), *tɔm* 'water' "is generally applicable to units, courses and bodies of water of all kinds and sizes." This effectively means that there are no words in Jahai that directly correspond to words like *river, creek, stream, lake* etc., although the language does allow to form complex terms on the basis of *tɔm* involving body-part metaphors to refer to specific aspects of *tɔm*s. What is important for the purpose of the present study is that, while there is not necessarily a set of lexical items whose semantics corresponds to, say, English words such as *mountain, river*, etc., there are lexical items that allow to *refer* to these configurations in the landscape, and it is not least the variation in the other semantic areas covered by the respective expressions that are an important aspect of the present study. In other words, cross-linguistic

comparison as carried out in the present study is necessarily based on extensional rather than intensional criteria (see Goddard and Wierzbicka 2010 for an overview of different approaches and the problems associated with them; one of the drawbacks of extensional analysis they mention is the difficulty of applying them to abstract semantic domains which are not investigated presently).

Note also that the above list does not exclusively feature meanings related to landscape, but also an array of other nature-related items. Many of the landscape terms and many of the other nature-related terms have a quite different ontological status. Some have relatively clear-cut boundaries (what Mark et al. 1999: 286 call 'bona-fide boundaries,' which "correspond to physical discontinuities in the world"), others don't. Related to this, some of them are attached and some are detached objects in terms of Gibson's (1979) ecological psychology: detached objects are wholes with clear-cut boundaries separating them from their respective environment and they can be grasped and moved by humans (this is also true of artifacts). The ontology of attached objects, on the other hand, is such that they do not have clear boundaries that would unambiguously delimitate them from their environment: they cannot be separated from the place where they are found, which, according to Smith and Mark (2003) is the case for landforms such as 'mountains' and 'valleys.' Generally, their conclusion is that "the pertinent basic level categories in this geographic domain of primary theory are precisely entities such as mountain, hill, island, lake, and so forth," which is one of the reasons why a number of them figure on the wordlist.

Other meanings on the wordlist, like 'wax,' denote masses, others, like 'nest,' denote individualized entities in the world. Some of the latter have component parts that themselves have lexical labels, others do not.[2] Some, like 'tree,' 'bird,' 'animal,' and 'plant,' participate in broader lexical taxonomies (in these particular cases, the life-form and unique beginner ranks as recognized in ethnobiology), while others do not seem to be embedded in taxonomic hierarchies at all, or at least not in a straightforward way. All these distinctions will allow for more fine-grained observations as to potential differences how these objects are treated linguistically in different languages.

It is clear that the environments of the speakers of the languages of the present study can differ quite drastically. While certainly the 'sun,' the 'moon,' and 'clouds' are experienced in all conceivable environments, maritime and hydrological concepts such as 'waterfall' and 'bay,' and possibly even 'rain' or 'river' are not necessarily. Thus terms denoting these entities may not be part of the vocabulary of, say, speakers of Ngaanyatjarra, who traditionally live in the Western Desert area of Australia. This statements probably also extends to concepts such as 'forest' or 'cave.' Similarly, New

[2] Compare the distinction between parts and pieces in Hayes (1985) and Cruse (1986: 157-60). "Parts" in this sense have "non-arbitrary boundaries and determinate function with respect to the whole" (Cruse 1986: 158-159; this notion draws close to the more traditional term of meronymy). "Pieces," in contrast are not characterized by these properties. For instance, the type bars and keys are parts of a typewriter but not pieces; when a typewriter is cut into several arbitrary portions by a hacksaw these are instances of pieces, but not of parts (example from Cruse 1986).

Guinea and Australia did not host any horned animals prior to their introduction by Europeans, which leads Laycock (1970: 1150) to state that this concept is "totally unsuitable for New Guinea" for elicitation. The non-universality of some of the concept on the meaning list is acknowledged.

3.2.1.2. Artifacts

1. airplane
2. ball
3. bed
4. belt
5. boat
6. car
7. chair
8. clock
9. glasses
10. house
11. key
12. knife
13. ladder
14. mirror
15. needle
16. paper
17. pen
18. rope
19. scissors
20. shoe
21. road/street/way
22. table
23. toilet
24. train
25. weapon
26. window

Artifacts on this list fall roughly into two categories: those that will be relatively recent arrivals in many areas of the world, introduced in the process of acculturation to (mostly) western-based societies, and those that are relatively basic tools (such as 'knife' and 'rope') that can be expected to have been present in most societies for a very considerable time span. There is thus a twofold division in the items in this semantic domain, and the purpose of this division is also twofold: first, it allows to investigate systematically whether there are differences with respect to motivation in languages with respect to the "old" artifacts and the newly acquired ones (it also allows to investigate related questions, such as whether lexically motivated terms for "old" artifacts entail a motivated lexicon for the recent cultural additions, etc.). More importantly, this list of artifacts allows to expand Brown's (1999) study on lexical acculturation in languages of the Americas, which reveals striking uniformities in the denomination strategies chosen, and thus will allow to answer the question as to whether these commonalities are peculiar to the Americas or are indeed recurrent on a global scale.

3.2.1.3. Body parts and body fluids

1. Adam's Apple
2. ankle
3. beard
4. bladder
5. blood
6. bone
7. brain
8. breast
20. rib
21. saliva/spittle
22. scar
23. skin
24. snot
25. sperm
26. stomach/belly
27. sweat

9. buttocks
10. calf
11. cheek
12. chin
13. eyeball
14. eyebrow
15. eyelash
16. nipple
17. nostrils
18. pupil
19. pus
28. tear
29. tendon/sinew
30. testicle
31. tongue
32. tooth
33. urine
34. uvula
35. vein
36. womb
37. wrinkle

Body-part terminology is the *locus classicus* of comparative lexicological and semantic research, and a huge amount of literature has amassed which explores the realm of body-parts cross-linguistically. The reasons for this are obvious: since everybody has a body (but not every language has a word for body, Wilkins 1996), this is indeed a semantic domain that is universal to human experience, and mereological divisions of some sort within this domain will be found in all languages. In addition, the human body is commonly assumed to be a fine example of the role that perceivable discontinuities play for the division of a whole into parts (see Enfield et al. 2006 for critical evaluation). Given the relatively large amount of research that has already accumulated on cross-linguistic categorization of the body (though not all claims that have been made are uncontroversial), the present study focuses on parts of the body that have received relatively little to no attention. Thus, 'eye' and 'nose,' for instance, are not included in the list above, although even these concepts would have had the potential to reveal interesting cross-linguistic facts, because they are - contrary to intuition - not always expressed by morphologically simplex items.[3] What is more, the present study extends the discussion to the somewhat more sensitive domain of terms for body-fluids, a semantic area where little is known about cross-linguistic denomination strategies.

[3] For instance, in the Polynesian outlier language Kapingamarangi 'eye' is *godo-mada* 'thing-see' (Lieber and Dikepa 1974, for diachronic data from other languages that are indicative of erstwhile morphological complexity see also Eilers 1973a).

3.2.1.4. Phases of the day and miscellanea

1. dawn
2. day
3. dusk
4. night
5. noon
6. sunrise
7. sunset
8. man (human being)
9. Saturday
10. virgin
11. widow

There is every reason to believe that humans in all parts of the world are aware of the interplay of day and night, and the transition phases between them. However, it is in fact an open question in how far this likely universality is reflected lexically. Also, although the phenomena in 1-7 of this list are clearly perceivable, they are, unlike most meanings in the semantic domains discussed so far, not tangible and do not have the typical "thinginess" associated with phenomenologically and temporally stable entities in the extralinguistic world. Insofar, this section of the wordlist is more experimental (and hence also much shorter) when compared with the other parts. The same is true of the miscellaneous meanings in 8-11.

3.2.2. HOW CAN COMPARABILITY BE ENSURED?

Meaningful cross-linguistic comparison presupposes that comparanda are kept constant on the level of the signified. If one looks up a particular word in dictionaries of two languages, how can one be sure that the translational equivalents found actually "mean the same thing"? Take, for example, the Greek word *límni* and Blackfoot *ómahksíkimi,* which both are glossed as 'lake.' How can one be justified in saying that Greek utilizes a simplex lexical item to denote 'lake,' whereas Blackfoot has a complex expression that consists of words meaning 'large' and 'water'? Translational equivalence does not automatically entail semantic equivalence (Behrens 2000). In spite of the strong intuition that when words in different languages are translational equivalents of each other, they share at least some common semantics, it would be quite naïve to assume a priori that this need always be the case for all investigated meanings. This is clearly articulated by Buck (1929: 216):

> Whoever deals with 'synonyms' has to face the fact that these are generally only roughly synonymous. Words from different languages do not often coincide in all their applications, they rarely cover quite the same ground. To treat every application separately, comparing words only in specific fully equivalent phrases, is a counsel of perfection which would so complicate matters as to wreck any comprehensive project.

In order to provisionally deal with this issue, and to avoid the worst undesired effects of the problem of cross-linguistic semantic comparison hinted at by Buck, the present study - as discussed earlier - is restricted to low-level referring expressions (what Lyons 1977 calls "first-order entities") on the assumption that expressions that refer to some entity in the real world are better suited for comparison than are terms denoting actions and events ("verbs"). First, the semantic properties of these referring expressions seem to be relatively stable and resistant to contextual modulation: Cruse (1986: 152) states that "[i]t appears to be a property of predicative terms such as verbs and adjectives that their meanings are context-dependent to a much greater extent than those of nouns." Note also Foley's (1997: 35) discussion of 'rock':

> *Rock* is a noun and describes, as nouns prototypically do, an object locatable in the physical world, in this case a hard, solid mineral object of the natural world. As rocks are concrete objects, they are freely apprehensible by our senses; we can see them, touch them, and, if they are small enough, manipulate them with our hands. Further, they are susceptible to changes in state: they can be moved, broken, crushed or thrown. Finally, the boundaries of what is a rock and what is not are fairly sharp, both from the surrounding space and from other objects. As we can manipulate a rock, we can determine its boundaries in space, and, by noting the features of this object bound in space in this way, we are able to sort rocks from trees or mothers. Given all these properties of the objects called rocks and human perceptual and cognitive mechanisms, it seems warranted to hypothesize that a noun corresponding to *rock* is rather a predetermined category in the vocabularies of all languages. This would not exclude further nouns in a language to refer to types of rocks such as *boulder*, *pebble*, and *stone*. These may or may not be found, but all languages would have the core term *rock*.

Foley goes on to contrast verbal meanings with those of prototypical nouns and notes that they are in many ways just the opposite: they do not have sharp boundaries that set them out perceptually against other entities in the world, etc.[4] Indeed, it is for instance hard to believe that the semantics of terms for 'sun' listed in dictionaries are semantically so incompatible to each other that cross-linguistic comparison would be rendered unjustifiable. Put differently, it is assumed that the core of the investigated meanings are, as elements of the phenomenological world, input to the human perceptual apparatus (see Quine 1973: 23 for similar argumentation in philosophy).[5] Semantically, they are assumed to be what Goddard (2001: 18) calls "approximate universals," that is, meanings that can be

[4] Note also that most of the meanings on the wordlist have the property of being relatively stable in time, a criterion used by Givón (1979, 1984) for the identification of nouns cross-linguistically.

[5] See Immler (1991: 40) for an earlier strong claim that certain meanings need to be expressed by languages of all human societies derived from "extrapolation from our implicit knowledge about our own language."

expressed in the majority of the world's languages. This does not presuppose or entail that there is a monosemous term exclusively devoted to any of the meanings, but rather, that there is some way in which languages allow to refer to those meanings, i.e. that there is an overlap in the extensional range encompassed by the lexical semantics of the compared lexical items.[6] In other words, as already alluded to above, the underlying conception of semantics is necessarily extensionalist rather than intensionalist. The variability that one can expect to find does not hinder investigation, but is rather among the research goals of the present study.

Further, anticipating some results of the present investigation, there is at the very least circumstantial *a posteriori* evidence justifying the assumption of cross-linguistic comparability of the investigated meanings. As will be seen later, the preferred way for languages to form complex terms for 'flame' is to conceive of them as being similar to a 'tongue' (Appendix E, 22), and, arguably, this points to a common perceptually-driven conception of 'flame.' Likewise, it is hard to believe that both Mali and Fijian should conceptualize 'flood' with reference to 'scraping,' if they did not share an essentially similar notion of 'flood.' This is circumstantial evidence only, and it cannot go full way in justifying the working hypothesis of at least rough semantic comparability, but it is reason for confidence that a comparative work such as the present one is not entirely misguided. Still, the general issue of semantic comparability and the justification for semantic analyses of the obtained data remains in spite of this first provisional measure. A more complete exposition of the issue and possible approaches to minimize problems caused by it follows in § 3.6.3.

3.2.3. THE WORDLIST AND THE NOUN/VERB-DISTINCTION

For the purpose of this study, nominal vocabulary is delimitated by the ability of the members of this class to refer (Searle 1969/1980, Lyons 1977). Nominals, in this sense, single out perceivable stimuli in the phenomenological world about which something can then be predicated. This study thus does something that is abhorred in grammatical descriptions of parts-of-speech systems in individual languages: it defines "nominal" vocabulary notionally, i.e. on semantic and pragmatic grounds, rather than on

[6] Compare Haspelmath (2007: 127): "Notice that for the purposes of typological comparison we do not need identity of strictly linguistic meanings. All we need is some level of meaning at which meanings must be commensurable." See, in addition, Kibrik (1986) for similar remarks, and Rijkhoff (2009) for critique.

morphosyntactic and distributional criteria (Schachter 1985).[7] This does of course not entail the suggestion to abandon the primacy of grammatical over semantic criteria in analysis of parts-of-speech systems. Rather, the present study investigated semantic configurations that are, on the conceptual level, *construed* as nominals (as "things" in Langacker's 1987a terminology) in the sense of Cognitive Linguistics (Langacker 1987b, Talmy 1988/2006) and that are therefore, at the linguistic level realized in the majority of languages by nouns (defined on language-specific criteria) rather than verbs. This is not always the case. To adduce an example from the database, the Oneida word for 'airplane,' *teka·tʌ́he?*, is, like other referring expressions in Oneida (Michelson 1990, Abbott 2000, see also § 4.6.4.3.2.), a straightforward verb on morphological grounds: it consists of the verb root *-tʌ-* 'to fly,' and the prefixes *te-* (called 'dualic' by Iroquoianists and having a variety of functions), *ka-* 'neuter agent' (the vowel of which is lengthened in the surface representation due to a phonological rule), and the habitual aspect suffix *-he?*. There is no nominalizing morphology whatsoever, and all bound morphemes occurring in the term are typically associated with verbs only. Thus, *teka·tʌ́he?* literally means 'it flies habitually,' and could probably be interpreted just in this way, i.e. predicatively, in an appropriate context (see Michelson 1990 specifically for these aspects of the lexicon of Oneida). Such phenomena are not restricted to North America, however. An even more extreme example from a different area of the world mentioned by Brugmann (1900/1981: 157) is Sanskrit *ki-vadanti* 'rumor,' which is literally translated 'what do they say?' In fact, it is a full-fledged sentence, with a finite verb form, and yet clearly capable of referring (compare also French *on-dit* 'rumor' from *on dit* 'one says).

For the context of the present study, it would make little sense to exclude terms like Oneida *teka·tʌ́he?* from the analysis on grounds of their formal realization as verbs. In fact, interesting cross-linguistic aspects of the lexical organization of the nominal domain would likely be lost by such a decision. In short, for the present study, it does not matter in practical terms whether the equivalent given for one of the meanings on the wordlist was stated in the consulted sources to be either noun or verb (or something else) in morphosyntactic terms.

3.2.4. THE WORDLIST AND THE NOTION OF BASIC VOCABULARY

Basic vocabulary has been defined along a variety of parameters, such as assumed cultural neutrality (Swadesh 1971), frequency of use (McCarthy 1999), resistance to borrowing and

[7] Note also that not all approaches to lexical and/or syntactic categories are willing to accept this fully and suggest instead to make use of semantic criteria, at least to some extent, in their delimitation, see e.g., from different perspectives, Lyons (1966) and Anderson (1991, 2011a: 66-67; 98-105).

morphological simplicity (Tadmor 2009, Tadmor et al. 2010), and others (see e.g. Huttar's 2003 discussion of the different points of time of emergence for certain meanings in the process of creolization and its relation to the notion of basic vocabulary). Strictly speaking, the present list cannot be claimed to be made up of vocabulary items that conform to any of these definitions. Many artifacts are certainly not culturally neutral (but, as noted above, they can be assumed to be familiar in most parts of the world by today, and the linguistic response in different parts of the world to these stimuli is the principal reason for their inclusion). Also, as already discussed, there are some topological concepts on the list that will probably not be familiar in some parts of the world. Likewise, some of the meanings will probably figure prominently in discourse across cultures, while others almost certainly won't. Finally, morphological complexity in the investigated items is the very notion that is to be determined in the present study, and therefore cross-linguistic realization by simplex terms would have been a poor candidate to serve as a guideline for the selection of meanings to begin with. However, there is a fair degree of overlap between the presently used wordlist and the nominal vocabulary in the Swadesh list and the Leipzig-Jakarta list established in Tadmor (2009) and Tadmor et al. (2010). Further, many researchers will have more intuition-based conceptions of what is and what is not basic vocabulary, and some of the concepts on the present list will not conform with intuitive judgments as to basicness. The present word-list is not intended to serve any other purpose than to be used for the present study. It is therefore best thought of simply as a list of meanings with some internal semantic diversification that seem worth investigating (some of which are basic in the sense that they figure on the pertinent lists) and for which, despite their referents' absence in some areas of the world, the majority of dictionaries can be expected to list equivalents.

3.3. SAMPLING

Cross-linguistic studies require some sort of language sample. To decide on the precise makeup of the sample is not an easy task. On the one hand, temporal and bibliographical restrictions need to be taken into account, on the other hand, the quality and validity of the generalizations that will be arrived at by evaluation of the sample data are a direct function of the soundness of the sample design.

The topic of the present work has never been studied in any systematic fashion from an explicitly cross-linguistic point of view, and it is not at all clear just what relevant factors might be at play in shaping the behavior of individual languages with respect to motivation in the nominal vocabulary. Of the several types of language sampling procedures available (see e.g. Rijkhoff and Bakker 1998, Bakker 2010 for an overview),

construction of a variety sample, as developed by Rijkhoff et al. (1993) and Rijkhoff and Bakker (1998), seems to be the most appropriate sampling procedure. This type of sample is designed specifically for use in exploratory studies where little is known in advance about the investigated variables, and it is known to produce highly reliable results (Widmann and Bakker 2006). A variety sample is constructed by measuring the internal diversity of language families in the language classification chosen to be underlying the sample. This is done by means of a fixed computational method called the Diversity Value technique. The Diversity Value is used to determine how many languages per family are to be included in the sample, given a predetermined desired sample size. Since it is impossible to sample fractions of languages, the smallest possible diversity sample logically must comprise at least one language per highest-order grouping (family) recognized in the underlying classification, with isolates and pidgin and creole languages each treated as one pseudo-family; the remaining number of languages are distributed over families as determined by the Diversity Value algorithm.

Having decided on the sample procedure, the underlying language classification and the desired sample size need to be decided on. Unfortunately, most extant word-wide classifications have their problems. Ruhlen (1991) operates with the assumption of a small number of macrofamilies largely determined by the contentious technique of mass lexical comparison. Voegelin and Voegelin (1977) are not always unambiguous in assigning a particular language to a particular family, and the Ethnologue's classification (Lewis 2009 and previous editions), while undisputedly the most inclusive of all available classifications in terms of the sheer number of languages, has the drawback that it is not at all made explicit how decisions as to subgrouping have been made, nor who made them. An alternative is Dryer (2005a), the classification employed in the World Atlas of Language Structures (Haspelmath et al. 2005). It is based on the Ethnologue's classification, but was revised in some places, and it is the work of a clearly identifiable author responsible for decisions. A further advantage of Dryer's classification is the relatively high number of families recognized in the Americas compared with other proposals, since the structural diversity of the linguistic situation in the Americas is argued to have not been sufficiently captured in previous typological studies by Dahl (2008). Dryer (2005a) was used as the classification to be used for the present study, although it has the drawback that it is relatively new and therefore the relevant sampling literature supplies no information as to precisely how remaining languages should be distributed over families. [8] Dryer (2005a)

[8] Note that the classification has been altered slightly in the subsequent online version of the World Atlas of Language Structures (Haspelmath et al. 2008) and may be subject to further refinements in the future. Here, the original classification as of 2005 if used. Further, it should be made explicit that choosing a particular classification

recognizes 93 highest-order groupings that contain more than one language, i.e. there is a minimum of 95 languages (plus one isolate and one pidgin or creole language) to be in principle included in the sample. This figure is already fairly close to Perkins's (1989: 312) recommendation of "[u]sing around a hundred languages for most linguistic samples to balance the requirements for representativeness and independence in samples." Unfortunately, as will be discussed later in greater detail, there is of course the well-known problem of bibliographical restrictions that prevent every single one of the 95 families from being represented in the sample, and the sample size would eventually drop significantly below the number of 95. Therefore, more than one language was included for some larger families, as when applying a Diversity Value algorithm like in the originally proposed design for a variety sample. Given that, as mentioned earlier, Dryer's (2005a) classification is fairly recent and could therefore be not considered in the original literature that developed the sampling technique to construct variety samples, it was necessary to devise an ad-hoc procedure to accommodate this sampling technique to Dryer's classification. One additional language was included in the sample for every 10th genus, which is a subordinate genealogical unit in Dryer (2005a), examples of which would be Romance or Germanic within Indo-European, that the family hosts. Thus Uralic, which has three distinct genera in Dryer (2005a), is represented in the sample by one language (namely Kildin Saami), Sino-Tibetan, having 14 distinct genera, is represented by two languages (namely Mandarin Chinese and Bwe Karen) and the large Niger-Congo family of Africa with as many as 25 separate genera is represented by the three languages Buli, Efik, and Mbum. Admittedly, this is a rather coarse mechanism, whereas the original method to compute Diversity Values and to determine how many languages to include per family is mathematically in contrast quite sophisticated. However, as a look at the appendices in Rijkhoff and Bakker (1998) reveals, the true virtue of the Diversity Value calculation plays out most strongly with very large samples, whereas in samples that are not significantly larger than the minimum sample size imposed by the classification used, the outcome is usually that one or two extra languages are added to the large families. Therefore, it is appropriate to be confident that the sample resulting from the admittedly coarse sampling method presently employed does not diverge drastically from one that would be constructed by thorough application of the Diversity Value technique, provided that, to repeat, sample size is rather small.

---

to stratify the sample does not necessarily entail that the typologist (nor of course, any of the authors of descriptive material used in his/her study) agrees with the classification in all cases.

On the basis of this technique, a genealogically balanced sample comprising 109 languages was arrived at, for which published materials were sought. Since the topic of this work is concerned with the lexicon, dictionaries were the prime target of this search, and indeed, availability of a dictionary was the principal guide when it came to the question as to which language(s) of the family were to be selected to represent it in the sample. In many cases, i.e. when lexical sources were readily available only for one language per family, this already determined unequivocally which language was to be selected for the sample. When several languages fulfilled this criterion, which is obviously the case for instance for the majority of Indo-European languages, but also for some minor families, for instance the Iroquoian languages, several of which come with excellent and detailed dictionaries, a language was chosen at random.[9] In addition, relevant grammatical information for the chosen languages was assembled either from the grammar sketch often provided in dictionaries themselves or, where this was not or not satisfactorily possible, additional sources for grammatical structure were consulted.

The aim of this study is (at least) two-pronged: while investigation of the quantitative aspect of lexical motivation is certainly a very important goal, qualitative semantic analyses as to recurrent conceptualization strategies and cross-linguistic differences in them are at least as relevant. However, the different questions pose different demands on the data: quantitative analysis requires a robust data representation for individual languages so the analysis can be considered meaningful. If only terms for a small fraction of the 160 concepts selected for investigation can be gathered from published sources, serious doubts arise as to whether they can be indicative for the language's profile with respect to lexical motivation. This is for instance the case for Rao of the Lower Sepik Ramu family of New Guinea, for which a mere 44 of the 160 concepts could be extracted from the consulted source. This is clearly insufficient for arriving at any conclusions beyond speculation as to the lexical profile with respect to motivation in Rao. On the other hand, the overall scarce representational score for Rao makes it no less interesting that the Rao word for 'cloud,' *grača*, apparently contains the word for 'sun,' *gra*, just like *núnik kás* 'cloud' in the much better documented Rama language of Nicaragua, contains *núnik* 'sun, day.' Given the different purposes that the sample needs to fulfill, it seemed appropriate to adopt the policy to use as much data as possible from different families for the qualitatively oriented part of this work in order to maximize the linguistic diversity the sample captures (after all, this is what it was designed to do!), but to introduce a threshold of 104, or 65 per cent, of meanings that need to be represented in the database so that a given language can be

[9] More precisely, a random number was generated at www.random.org and the language whose position in an alphabetical list of the languages in the family corresponded with the generated random number was chosen.

said to be represented adequately to give a reasonable picture of its behavior with respect to motivation in the nominal vocabulary. Consequently, the sample is split for the different purposes. Languages with more than 65 per cent of lexical items represented are grouped in what will be called the CORE SAMPLE. This sample comprises 94 languages. Since statistical analysis, which will be carried out throughout in the search for typological correlations, requires datapoints (in this case: languages) to be independent of one another, there is a restricted STATISTICS sample of 78 languages which features only one language per family recognized in Dryer (2005a). At times, statistics will also be performed on a more extensive sample, with appropriate counter-measures taken to ensure validity of the performed tests. Unless specifically indicated, all statistics is carried out on the basis of the statistics sample (moreover, there is the STATISTICS VALIDATION SAMPLE, the purpose of which is described in § 5.4.1.). Together with the 18 languages that did not fulfill the 65 per cent criterion, the core sample constitutes the EXTENDED SAMPLE 1 (EXT-1 for short). This sample still has the property of being genealogically balanced. In addition, in the course of preparation, in some cases materials for additional languages were gathered, for a variety of reasons. First, it was sometimes the case that on the basis of the originally chosen language the 65 per cent threshold could, contrary to what was expected from the general appearance of the source, not be attained. If another language from the same family had a dictionary available, data was gathered for that language, and quite often, more data could be found and this language could be included in the core sample. Second, in the course of the work a number of sampled languages chosen behaved contrary to what was expected in one aspect or another. Of course, these languages were not removed from the sample (this would be deliberate manipulation), but in some of these cases an additional dataset was, where easily available, gathered for a closely related language to be able to ascertain whether there is reason to believe that the first language was simply aberrant in its behavior. Thirdly, publication of the World Loanword Database (Haspelmath and Tadmor 2009c) made available a set of high quality vocabularies that have the particular advantage that they explicitly include information about the morphological analyzability of the vocabulary items listed, which is an invaluable asset for a study such as the present one. Unfortunately, the World Loanword Database was published only when the database for the present project was already in a fairly advanced state, so that not all potentially relevant data could be featured in the core sample. Still, this opportunity was taken to augment the sample further at relatively little cost by including data from six languages in addition to those that are already featured in the core sample. Together with the EXTENDED SAMPLE 1, the abovementioned data constitute the EXTENDED SAMPLE 2 (EXT-2 for short), which comprises data for a total of 148 languages. This sample is not genealogically balanced, and must be considered a mere convenience sample in terms of Rijkhoff and Bakker (1998). The EXT-2

sample is used as the basis for the extensive list of recurrent lexico-semantic associations found in Appendix E and for many evaluations with regard to semantic associations in chapter 6. A list of all sampled languages and consulted sources is in Appendix A. Figure 1 is a map of the world showing the location and sample affinity of all languages investigated.

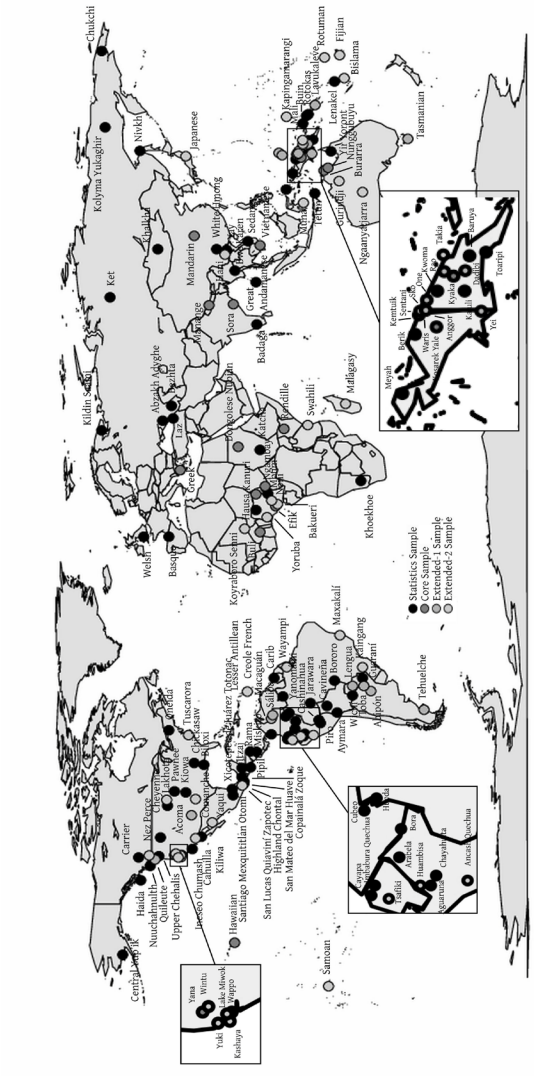


fig. 1: languages in the different samples and their location

3.4. SOME BASIC ASSUMPTONS AND THEORETICALPREREQUISITES

3.4.1. WHAT'S IN THE LEXICON?[10]

Central to a study in lexicology is to make clear from the outset the underlying conception of the lexicon. According to a very widespread view that has a long tradition in linguistics, the lexicon is the "locus of unpredictability of form-meaning associations" (Anderson 1985: 4). This view has received backup from early cognitive studies on the storage capacity of the human brain. In line with the then prevailing mind-as-machine-metaphor (see Daugham 2001 for review), the human brain seeks to maximize storage efficiency: therefore, only idiomatic expressions (simplex lexical items which are by definition idiomatic because of the arbitrariness of the relation of meaning and form, idiomatic compounds and idioms themselves) are assumed to be represented by an entry in the lexicon of their own, whereas all non-idiomatic expressions do not have an entry, because their meaning can be derived by productive rules. As DiScullio and Williams (1987: 3) famously summarize, "[t]he lexicon is like a prison - it contains only the lawless, and the only thing that its inmates have in common is lawlessness." However, conceptions of the lexicon differ quite radically depending on the point of view and purpose of the analyst. Pawley (1996a) and Himmelmann (2004) distinguish between the "grammarian's lexicon" and the "lexicographer's lexicon" and show that assumptions about what "is in the lexicon" differ quite radically in the conceptions of the two research traditions. In the words of Pawley (1996a: 189):

> Whereas a lexicographer's lexicon is essentially about conventional expressions for conventional concepts, a grammarian's lexicon treats exceptions to well-formedness constraints in grammars. There is a common membership in the ideal lexicons of lexicographer and grammarian, namely, the form-meaning units that are either unanalyseable or irregularly formed. Where their memberships diverge is in the treatment of complex expressions that are both well-formed and more or less conventional. In the languages that I am familiar with, the class of such expressions (derived and compound words, catch phrases, speech formulas, etc.) is indefinitely large. ... But it is probably safe to say that, for a given language, the sort of minimalist lexicon posited as the ideal in certain grammarians' models will contain less than ten percent of the lexemes that might be included in a lexicographer's ideal comprehensive dictionary.

For the present work, the lexicon of languages is the topic of investigation in its entirety. This entails that it is taken to be the repository for any conventional form-meaning

[10] The title of this section owes its wording to Jackendoff (2002).

pairings that are memorized by speakers as the normal way to refer to a certain extra-linguistic entity (see also Schultze-Berndt 2000 for a similar position and Bybee 1995 for arguments from the perspective of word frequency). These include morphologically simplex items, compounds, derivatives, phrasal lexemes, idioms, etc. Conventionality, however hard to pin down in a theoretically principled fashion (see Pawley 1986 for some proposals), rather than theoretical considerations from the point of view of a particular linguistic theory, are the characteristic taken to be defining feature of whether a unit is in the lexicon or not.

3.4.2. COMPOSITIONALITY OF MEANING

Closely connected to the conception of the lexicon is the question if, and if yes to what degree, one accepts the existence of the principle of compositionality of meaning, which states, roughly, that the meaning of a complex expression is the sum of the meaning of its parts and the combinatorial principles that link the parts together. Frege (1892) introduced this principle initially with reference to sentences, but its scope was later expanded to the word-level as well. Thus, it is said that compounds such as *kitchen knife* are amenable to an interpretation in line with the principle of compositionality of meaning (a 'kitchen knife' is a 'knife' that has got something to do with the 'kitchen,' and is most likely used for preparing food). Other types of compounds, such as *hot dog*, along with phrasal idioms such as *kick the bucket*, are said to be not analyzable compositionally: a 'hot dog' has got nothing to do with 'dogs,' and 'to die,' the meaning of *to kick the bucket,* has got nothing to do with either 'kicking' or 'buckets' altogether. These cases appear straightforward, but often it proves to be much more difficult to arrive at a well-founded statement whether a given lexeme or construction is interpretable compositionally. Here, the precise delimitation of what an interpretation in line with the principle of compositionality is becomes intuitively less clear. Often the criterion of purported additional semantic features is invoked to argue that a given lexeme is non-compositional semantically and must therefore be assumed to be stored in the lexicon. For instance, the addition of a semantic feature of 'professionally' or 'habitually' is taken to be evidence to show that the English agent noun *teacher* is not, or not fully, interpretable compositionally: a 'teacher' typically is not anybody who happens to teach, but someone who does it professionally in a school or in another educational institution (Kastovsky 1982). This criterion is, though, vague in that "it is not clear what the 'addition of semantic information' should be taken to include" (Bauer 1983: 57). Intuitively, one would be inclined to say that *teacher* is more readily interpretable compositionally than classic instances of idioms such as *to kick the bucket*, and indeed, compositionality of meaning is probably best seen as a scalar phenomenon rather than an absolute dichotomy. As summarized by Hoeksema (2000: 856), "[t]here is a gliding scale from fully compositional

(or motivated) to fully idiomatic (or conventional)," and the principle of compositionality is thus "not an empirical constraint on possible interpretations, because it may be violated, as it is in the case of idioms." Blank (2001) goes one step further by arguing that *any* morphologically complex lexeme can be said to be non-compositional semantically and that, once a neologism becomes established as the current name of an entity, there are always slight but notable deviations from the compositionality principle.

The principle of compositionality has played a central role in many approaches to the lexicon: if a given unit is not amenable to compositional interpretation, it is said to be 'lexicalized,' in line with the grammarian's conception of the lexicon as a storage of units with arbitrary form-meaning association. This class then comprises simplex terms (which are idiomatic by the very nature of the arbitrariness of the linguistic sign) and non-compositional complex expressions (Herbermann 2002). However, the principle of compositionality is not universally accepted, at least in its strong form, by researchers in all approaches to semantics and lexicology. General critique on its validity comes, in addition to the problems mentioned above, from a variety of approaches. Cognitive Linguists, for instance, have argued forcefully against the cognitive reality of the principle of compositionality. For example, Taylor (2002: 550) states that in actual utterances rather than in the semantics of isolated words "[s]trict compositionality is rarely, if ever, encountered." Likewise, Kövecses (2006: 325) suggests "that idiomaticity (as opposed to regularity and predictability) is the default feature of natural languages."

What is more, while the principle of compositionality may be a useful heuristic for a semantic classification of complex expressions, recent psycholinguistic research on compound processing casts doubt on its psychological reality as well. Libben (2003) reports priming effects for both members of semantically opaque (i.e. non-compositional) compounds, and Semenza and Mendini (2006) show that decomposition in lexical processing occurs irrespective of whether they would be classified as compositional or non-compositional. According to Libben (2006: 6), "[i]f a compound word has been presented often enough so that it can be lexicalized, it is stored as a representation that can be retrieved as a whole. This, however, does not shut down the process of morphological decomposition for that word, nor does it sever links between the whole compound word and its constituent morphemes." In other words, so-called idiomatic compounds are not, as one might assume, simply stored as holistic units disregarding their morphological make-ups.

The conclusion is that compounds, and presumably also other complex expressions, are not necessarily either stored in the lexicon or derived productively, but that it is possible for them to be stored in the lexicon as a whole and *at the same time* to be decomposable in their constituents. Further, Weldon (1991) presents evidence that

processing a compound facilitates subsequent recognition of one of its constituents and Inhoff et al. (1996) also find evidence for the opposite direction. Both results support a view of a mental representation of complex expressions that is linked to their constituent parts. Similar views seem to trickle down to more progressively-minded circles in Generative Linguistics: Jackendoff (2002: 30) says that "there is no reason to believe that the brain stores information without redundancy" and allows, in his theory, "the possibility of stored words with regular morphology, which, since they can be generated by free combination, are totally redundant" (2002: 44). This view is in line with Cognitive Grammar (see e.g. Langacker 1987a: 28).

Summarizing, recent research at least opens up the possibility that compositionality of meaning, at least when applied to the word-level, is an artifact that is constructed theoretically and *a posteriori* from a cognitively and psycholinguistically already preexisting word meaning rather than an actually meaningful characteristic of linguistic units. These doubts are another reason for not operating with a conception of the lexicon that regards it as a redundancy-free storage of units with unpredictable form-meaning pairings, and any purely semantically grounded distinction between compositional and non-compositional complex words will play no role in the further discussion. Instead, the lexicon is understood as the sum of conventionalized expressions that are available to speakers to say something about entities and events in the extralinguistic world, regardless of whether these are interpretable compositionally (by any definition) or not.

3.5. LEXICAL MOTIVATION DEFINED

As alluded to in chapter 1, where the topics under investigation were introduced, as far as the semantic side of lexical motivation is concerned, there is evidence that complex terms cannot fruitfully be separated from and studied without also considering semantic patterns that do not involve morphological complexity. Consider the following terms for 'tree bark' from Bezhta and Mbum.

(1.) a. Bezhta *beš* 'skin, bark'
b. Mbum *ŋgàŋ-kpù* 'skin-trunk/tree' ='bark'

In both cases, a presumably identical or at the very least highly similar process is taking place which links the meanings 'skin' and 'bark' in some as yet to be established fashion. From a semantic point of view, this is the important commonality between the terms in the two languages. The difference lies in the formal realization: in Bezhta no overt sign of the relationship on the level of the signifier is present (*beš* can refer to both 'skin' and 'bark'),

whereas in Mbum, the meaning 'bark' is realized by a morphologically complex term. But, to reiterate, the semantic pattern is fundamentally the same. Ungerer (2007: 652) correctly points out that "[j]ust like additional meanings of simplex lexical items, word-formation items can be understood as encoding extensions. ... The only difference between simplex and word-formation items is that in the latter, additional meaningful components, both lexical items and affixes, are added" (see also Apresjan 1974: 17, Willems 1983: 426, Evans 1992: 478).[11] Then, the question arises whether it is appropriate to view (1b.) as interesting because of its morphological complexity and dismiss structures such as the one in Bezhta from consideration of semantic comparison. The point of view adopted here is that this would be not only implausible from a semantic point of view, but that it would also potentially distort the results that would be obtained from a study such as the present one. After all, it is possible that (1a.) is a pattern favored by some languages while (1b.), involving overt marking, is a pattern preferred by others, and potentially interesting typological observations could be obtained from a systematic evaluation of such differences. Examples like those in (1.) could be multiplied almost ad infinitum on the basis of the collected data:

(2.) a. Oneida *onékliʔ* 'grass, hay, straw'
b. Hawaiian *mauʻu maloʻo* 'grass dry' = 'hay, straw'

(3.) a. Hausa *ido* 'eye, spring of water'
b. Meyah *mei eitéij* 'water eye' = 'spring'

(4.) a. Upper Chehalis *páx*^{w}- 'smoke, steam'
b. Highland Chontal *liguxís gajah* /liguxís lajah/ 'smoke water' = 'steam'

(5.) a. Ngaanyatjarra *kantja* 'chin, beard'
b. Bwe Karen *khɛ ʃʊ* 'chin hair' = 'beard'

Drawing a strong line between realization of a given lexico-semantic association by "polysemy" and by morphologically complex items therefore is somewhat reminiscent of drawing an arbitrary line in the sand. It also does not make sense in the light of the

[11] Colman and Anderson (2004) also recognize this relationship, and moreover argue that metonymies, because of being systematic in some instances (for instance the regular pattern of using the name of a garment to refer to the person wearing that garment) ought to be treated on a par with conversion, traditionally conceived of as derivation (and hence word-formation).

frequently encountered flux between complex terms and polysemy by redundant marking (Koch 2008: 125-127), as when the Maxakalí word *gõy* ranges semantically over 'smoke,' 'cloud,' and 'fog' and there are compounds with *hãm* (reduced from *hahãm* 'land') and *tex* (reduced from *tehex* 'rain') to single out the meaning 'fog' specifically (see Urban 2011 for a diachronic interpretation of such seemingly redundant marking). The close ties between formal realization by analyzable terms and by "polysemous" terms has been recognized in the framework developed by Koch (2001), which is applied for instance in Blank (2003) to diachronic questions and which is elaborated on in Koch and Marzo (2007). It is the most systematized framework to the present date to describe lexico-semantic relations, and the approach will therefore be discussed in some greater detail, as it constitutes the baseline against which the present account will be developed. From now on, this framework will be referred to as the "Koch approach" for convenience. It operates with a redefined definition of motivation, a term popularized by Saussure (1916/1967). In the Koch approach, lexical motivation, which in this narrow sense excludes onomatopoeia as a motivating device, is defined as follows, and essentially this is also the definition adopted here, albeit, as will become clear shortly, its elaboration differs somewhat:

> A lexical item L_1 is motivated with respect to a lexical item L_2, if there is a cognitively relevant relation between the concept C_1 expressed by L_1 and the concept C_2 expressed by L_2 and if this cognitive relation is paralleled by a perceptible formal relation between the signifiers of L_1 and L_2 (Koch and Marzo 2007: 263).

This is a rather technical definition, and it is therefore instructive to provide a diagrammatic representation of the underlying idea. This can be done by means of a "motivational square" (Koch and Marzo 2007: 264). The general schema is seen in figure 2.

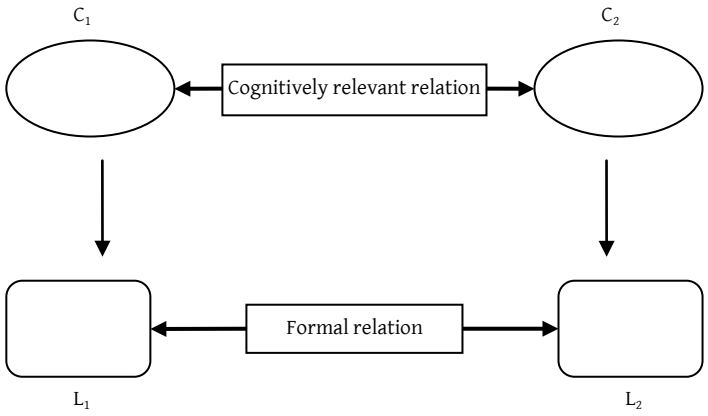


fig. 2: motivational square, redrawn from Koch and Marzo (2007: 264)

Different subtypes of formal and semantic relations (these are generally called "cognitive relations" in the Koch approach; here, it is preferred to use "cognitive" somewhat more cautiously) can be shown diagrammatically by specifying the formal and semantic relations that hold in a particular lexical item. Koch and Marzo (2007) use the concept 'pear tree' as a general example. In French, the word for 'pear tree' is *poirier*, which is formed by suffixation of *-ier* to *poire* 'pear,' shown diagrammatically in figure 3.

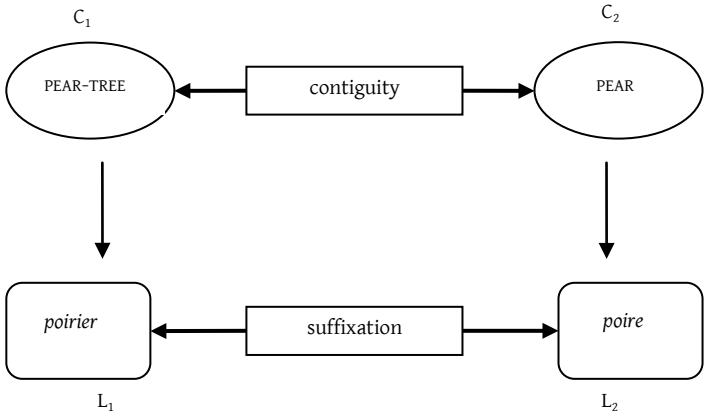


fig. 3: motivational square for a derived term, redrawn from Koch and Marzo (2007: 264)

In Hungarian, the word for 'pear tree,' *körtefa*, is a compound, consisting of *körte* 'pear' and *fa* 'tree.' For such cases, the Koch approach proposes a quite complex modification of the motivational square that seeks to account for both members of the compound, see figure 4. Essentially, this diagram is to be read as proposing a semantic relationship of taxonomic subordination between the concepts 'pear tree' and 'tree' on the one hand (a 'pear tree' is a kind of 'tree'), and a semantic relation of contiguity between 'pear' and 'pear tree' on the other hand. These semantic relations are mirrored by the formal make-up of Hungarian *körtefa*. It is at least questionable whether such a complex account is really necessary, for both theoretical and practical reasons, and it is here that the present approach will depart most pronouncedly from the Koch approach. The cognitive relation that is said to hold between the concepts 'pear' and 'pear tree' is that of contiguity, "the fundamental connection underlying all kinds of frames, scenarios, scripts etc. and including part-whole relations" (Koch and Marzo 2007: 262, for the related notion of "engynomy," see Koch 2001: 1144-1145).

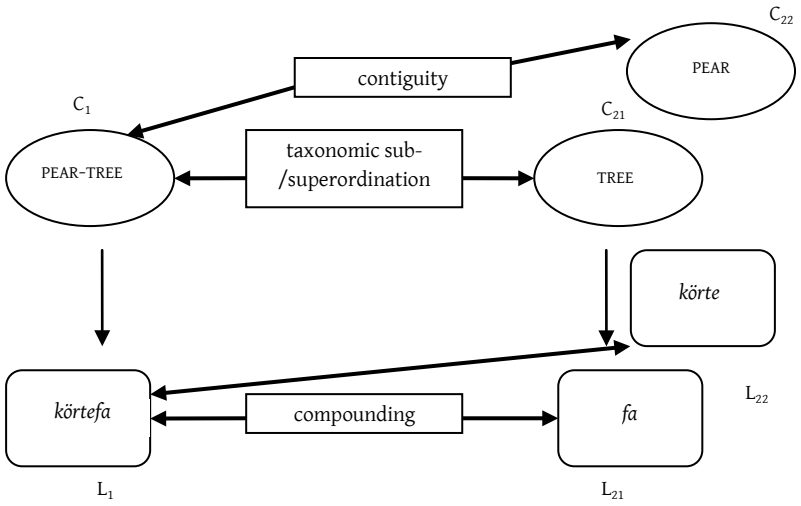


fig. 4: motivational square for a compound, redrawn from Koch and Marzo (2007: 266)

The same semantic relation is also present in Russian, which expresses both 'pear' and 'pear tree' by means of a single lexical item, *gruša*, according to Koch and Marzo (2007).[12]

[12] Note, though, that some native speakers of Russian are unwilling to accept this, preferring the complex structure *gruševoe derevo* for 'pear tree.'

Thus, the relevant formal relation is simply formal identity, yielding a reduced "motivational triangle" (figure 5).

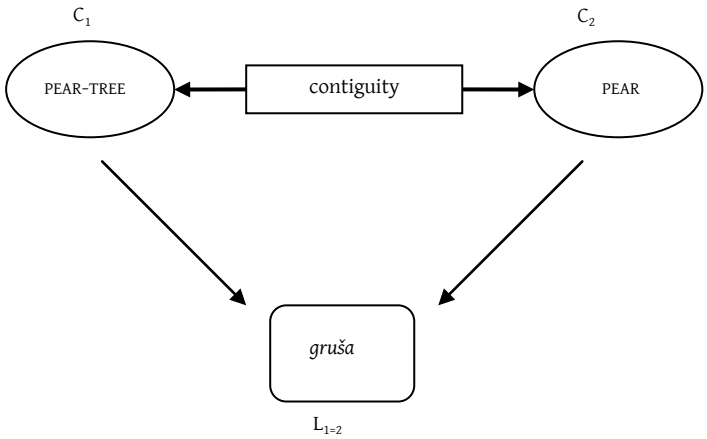


fig. 5: reduced motivational square for a polysemous term, redrawn from Koch and Marzo (2007: 265)

Finally, it is often the case that there is a semantic relationship between two concepts but no formal relation or vice versa (compare Anderson 2011b: 282 from the perspective of Notional Grammar). These are by definition not cases of lexical motivation in the sense used here, as for this to be diagnosed, crucially, both need to be present. Indeed, to use examples from Koch and Marzo (2007), it is intuitively clear that there is a semantic relationship between 'vines' and 'grapes' (one of "contiguity"), but in French, the two concepts are expressed by formally completely unrelated lexical items, *vigne* and *raisin* respectively. On the other hand, there can be a clear formal relation on the level of the linguistic item, but no clear semantic relation. If this formal relation is formal identity, this is the situation commonly known as homonymy.[13]

[13] Koch and Marzo (2007) apply this classificatory grid to the 500 most frequent lexical items of French, arriving at the conclusion that 65% of items in this set are motivated in their sense. Two thirds of these cases are due to polysemy; Koch and Marzo (2007: 278) attribute this to the typical situation of high-frequency vocabulary being highly polysemous. They suggest that such high rates would drop if less frequent and more specialized vocabulary items were to be analyzed, while cases of overt motivation would increase. Thus, the inclusion of polysemy (which is laden with the intricate problems of identifying the direction of semantic extension and highly specific secondary meanings, which are dealt with in a rather arbitrary fashion by Koch and Marzo 2007: 281-282) significantly

While in the Koch approach, the inventory of formal relations is sensibly open to be able to account for the different devices of word-formation that are found cross-linguistically, it is proposed that there is a "universal, closed inventory of seven cognitive relations" (Koch and Marzo 2007: 269). These are conceptual identity, contiguity, metaphorical similarity, co-taxonomic similarity, taxonomic superordination, taxonomic subordination, and conceptual contrast. In fact, Koch and Marzo (2007: 273) make quite strong claims about the nature of semantic relations: "The inventory of cognitive relations is universal and language-independent. As shown by phenomenological and gestalt psychological research as well as by historical semantics and as corroborated by empirical studies ..., there simply exist no other relations to connect conceptual contents." This view can be challenged (for instance, neither Gentner 1993 nor Levinson 1994 rely on the notion of metaphor), and Chaffin (1992: 254) points out that "[t]he widespread use of relations to explain other phenomena has often led theorists to treat semantic relations as if they were thoroughly understood. This is not the case." Although the notions of metaphor and metonymy figure in the present framework (albeit in a modified version), no claim shall be made that the inventory of semantic relations to be developed is universal, nor that it is exhaustive, nor even that it is indeed the best possible inventory. The only claim made is that the inventory of relations is a workable one from a cross-linguistic point of view.

Another question is whether it is justified or indeed feasible to take an reductionist approach to semantic relations postulating a closed set of cognitive principles in the light of highly culture-dependent patterns of semantic associations. Therefore, the present inventory, characterized by a fundamental distinction between contiguity and similarity, is merely meant to offer a useful categorization on some level of semantic detail – more coarse-grained in the case of this fundamental distinction, more fine-grained in the case of the subtypes to be distinguished. It would of course be possible to make even more-fine grained distinctions here, potentially ad infinitum, until every relationship between the senses of a given word or constituents of a complex term is assigned its own type of semantic relation.

---

contributes to the overall outcome of their investigation. As for the semantic dimension, they report that in 57.6% of the cases the relation is one of contiguity, 18.6% are characterized by metaphorical similarity, and about 10% are diagnosed as taxonomic in nature.

3.6. A TYPOLOGY OF FORM-MEANING-PAIRINGS

Lexical motivation as outlined above is a cover term for morphologically complex and polysemous lexical items, whose structure on the level of the linguistic sign is mirrored by corresponding semantic transfers. The goal of this section is to develop a typological grid for the classification of form-meaning pairings that can occur in motivated linguistic items. According to common lore, linguistic signs, like signs in any other semiotic system, are (at least) two-sided entities: they consist of a string of sounds and a meaning, typically arbitrarily, associated with it. Therefore, a classification of form-meaning pairings will necessarily involve a classification of the formal properties of the signifiers as well as a system of describing the semantic relations that hold between the constituent parts of a complex expression or the different meanings associated with a single polysemous lexeme (as recognized in the definition of lexical motivation in the Koch approach). It is the interplay of these factors which promises to yield interesting typological generalizations.

3.6.1. A CLASSIFICATION OF FORMAL RELATIONS

To simplify the discussion of formal relations, it is helpful to keep the semantics and the semantic processes (which will be dealt with later) of compared terms constant as far as possible. Consider the terms for 'flame' in (6.), some of which are characterized by some sort of lexical connection of 'flame' with 'tongue':

(6.) a. Hausa *harshe* 'tongue, flame'
b. Khoekhoe *ǂnora-b* 'flame-3SG.MASC
c. Lenakel *nam-nam-* 'tongue-RED-'[14]
d. San Mateo del Mar Huave *netitit* /ne-atitit/ 'NMLZ-to.flame.much'
e. Toaripi *a-uri* 'fire-tongue'
f. Kildin Saami *tōlnjūxxčem'* /tōl-ɲūxxčem'/ 'fire\GEN.SG-tongue'
g. Swahili *ulimi wa moto* 'tongue of fire'
h. Fijian *yame(yame) ni buka* 'tongue(-RED) POSS fire'
i. Rama *abúng ngárkali ~ abúng ngarkalima* 'fire flame'

Example (6a.), from Hausa, cannot be broken down into meaningful parts, i.e. morphemes. However, its semantic structure is such that there are two meanings, 'tongue' and 'flame,' associated with it.[15] Such a situation is commonly referred to as POLYSEMY, that is, one

[14] Considered "possible" in the consulted source.

[15] It may seem odd that a classification of formal properties of linguistic items should make reference to semantic properties. But, as Wälchli (2005: 136) observes, "since semantic classifications concern linguistic entities that have

lexical item with two or more related but distinct senses. Examples like (6a.) will be said, following François (2008), to be instances of COLEXIFICATION of two different senses which are expressed by the same lexeme; discussion as to why to avoid the more traditional notion of polysemy will follow.[16] Colexification as a technical term includes the special case of heterosemy (Lichtenberk 1991, Enfield 2006), in which "the different but related meanings of a given morpheme are associated with distinct grammatical contexts" (Enfield 2006: 297), and no attempt is made to distinguish heterosemy as a special subtype in the analysis. Examples (6b. – 6i.), in contrast, are morphologically complex or ANALYZABLE. ANALYZABILITY will be used as a cover term to characterize morphologically complex expressions formed by any language-specific syntactic or morphological process of word-formation. Analyzable term (or analyzable lexeme, lexical item, word) therefore will be used interchangeably with morphologically complex term (lexeme, lexical item, word) throughout the following chapters. This term is chosen to reflect that it is the result of the linguist's analysis of the data, and TRANSPARENCY will occasionally be used to refer to the ability of a speaker, when contemplating a particular linguistic item in an act of meta-linguistic reflection, to decompose it and identify its meaningful parts. The two notions will not always, perhaps in a considerable number of cases, coincide: as any field linguist will have experienced, the morphological make-up of a complex term, whose constituents may be perfectly clear to a linguist even at first sight, may not actually be perceived by speakers in the same way.[17] For instance, Kashaya *hasime* 'beard' is readily analyzable by the linguist as consisting underlyingly of *ʔaha* 'mouth' and *sime* 'fur' (the former shortened by the regular and frequent phonological process of aphesis, Oswalt 1961), but when asked, a consultant was entirely unaware of the morphological structure. Clearly, as suggested by Ullmann (1962: 93), transparency as opposed to analyzability would be an appropriate, or perhaps even more appropriate, level of analysis as well, but it is hard to operationalize in a consistent manner, and too time consuming to allow large-scale cross-linguistic investigations. As a further complication, even complex terms formed by synchronically non-productive patterns of word-formation may occasionally be transparent in this sense to speakers,

---

both semantic and formal components, semantic classifications of linguistic phenomena are never purely semantic. Thus, a classification of compounds cannot do without the notions of parts and wholes of compounds, which remain basically formal notions, even if only their meanings are considered."

[16] Compare also Gil's (2004) notion of macrofunctionality.

[17] As Rice (2012: 70) puts it, "post-hoc analyzability by the linguist does not always spell automatic conceptual recognition on the part of native speakers ..."

while perfectly regular complex expressions formed according to synchronically productive rules may not be (see Wilbur and Menn 1975 for discussion).

The Lenakel term for 'flame' in (1c.) is analyzable in the sense just outlined. More precisely, it is formed from *nam-* 'tongue' by means of reduplication. This is a case of DERIVATION, a process of word-formation that creates new lexemes by means of the application of some morphological operation to a lexical root, in this case reduplication. Characteristic for derivation, then, is one lexical element and one non-lexical derivational element. At least for the present study, complex terms formed on the basis of a lexical root by addition of a nominal classifier in languages that feature them (such as Bora, where the word for 'cave,' *ííñuhéju*, consists of lexical root *íñu* 'earth' and the classifier for hole-like objects, *-héju*), are subsumed under the label derivation. Consequently, (6c.) is called DERIVED.

Example (6d.) from San Mateo del Mar Huave is morphologically complex by virtue of being derived from a verb with a nominalizing suffix. Such processes serve to convert a lexical root from one lexical category to another, without adding anything to the semantic content of the lexical root (obviously, this statement pertains exclusively to nominalizers such as that in 6d. and not, for instance, to locative and instrument nominalizations). Examples such as these are clearly derived, but of a particular type that will be called PLAIN to highlight the semantic neutrality of the process in question. Plain derivatives cannot, according to the definition, be called motivated in the same sense as other derivatives can, since they are semantically neutral and do not establish lexical connections in the same way as other complex lexical items do. Consequently, lexical items of this type are not taken into account when the amount of motivated terms in each language is assessed quantitatively.

(6e. - 6i.) are examples of languages in which the term for 'flame' contains two morphemes with LEXICAL meaning: 'fire' and 'tongue' in (6e. - 6h.), and 'flame' and 'fire' in (6i.). Further subclassification, at least theoretically, is possible into COMPOUNDS and PHRASAL LEXEMES (which, as discussed above, should be allowed as inhabitants of the lexicon in the present approach): compounds would be complex lexemes formed according to morphological rules of word-formation by the combination of two preexisting lexemes (see Bauer 2009 for a concise overview of properties of compounds in a variety of languages), while phrasal lexemes, although conventionalized expressions with a fixed meaning in the respective language, follow the respective rules of syntax. Gravelle (1998: 569-570) makes clear that noun-noun combinations in Meyah obey morphological rules and can be identified by certain typical properties. Unfortunately, however, such explicit statements on the status of complex terms are not available for the majority of the sample languages, with the result that it is often impossible to distinguish the two types based on the

literature. (6e.), from Toaripi, looks at first glance like a typical compound: the term consists of two lexical morphemes with no additional morphology that links the two morphemes in any way. In contrast, (6f. - 6h.) might be hypothesized to be phrasal units, on grounds of the presence of typically phrasal morphology: in (6f.), from Kildin Saami, the dependent *tōl* appears in genitive case ("internal inflection" in terms of Bauer 2009: 346-7), and in (6g.) and (6h.) there is a free-standing morpheme that links together the two lexical elements of the form. Unfortunately, to say on the basis of these observations that (6e.) is a compound and (6f. - 6h.) are phrasal lexemes is based on mere eyeballing, does not rely on clear-cut criteria, and is therefore not legitimate.

There are several criteria available to distinguish between compounds and phrasal units that must, however, be applied carefully on a language-specific level (Aikhenvald 2007: 24-28). One is the criterion of separability: phrases, as syntactic units, are expected to be separable straightforwardly by additional modifying elements such as adjectives, whereas compounds are not. Haspelmath (2002: 158) mentions that in Hausa, compounds "clearly resemble possessive constructions in that they show head-dependent order and a relation marker" and that "[t]here are no phonological or morphological properties that would distinguish such compounds from possessive phrases..." Clearly, the Hausa evidence presented by Haspelmath militates once more against making any judgements based on superficial inspection. Haspelmath goes on to state that when adjectives are inserted to modify the respective constructions, "it becomes clear that the compound is inseperable, whereas the phrase is seperable" by an intruding adjective. This is a consistent and applicable criterion for Hausa, but it is a comparably subtle one. Also, it is certainly not applicable for all languages, for which different criteria must be invoked (see Matthews 1974: 94-100 for discussion of the situation in English). But even when applied on a language-specific level, criteria often fail to establish a waterproof distinction. For instance, Schütz (1985: 37) refers to complex Fijian lexemes with the structure root *ni* root as in (6h.) as compounds. Confusingly, though, he calls the sequence of *ni* and the second root "the *ni* phrase" and subsequent discussion (1985: 451-455) makes clear that indeed *ni* is best considered a preposition that projects a prepositional phrase indicating a general meaning of possession or association, in other words, that structures as (6h.) indeed can be considered phrasal. Schütz (1985: 38) also notes that "[f]or Fijian, there are no phonological clues to distinguish between a phrase and a word." Further, for Toaripi, Brown (1972: 206) notes that "[i]n compound forms, i.e. words composed of only free forms, the closeness of the juncture varies," which raises at least doubts about the status of so-called compounds in Toaripi. Summing up, distinguishing between compounds and conventionalized phrasal units is a matter of careful language-specific analysis and cannot be made on the basis of cursory inspection. Therefore compounds and phrasal units are not distinguished here, and

LEXICAL ANALYZABLITY is used as a deliberately ambiguous cover term to characterize both kinds of lexical units, compounds and phrases, based on more than one lexical stem. Importantly, unlike in some other studies, compounding as a type of word-formation is subsumed here under the lexical type of analyzable items, and thus not terminologically subsumed under derivation.

Further, differences in the presence of additional grammatical material that different languages require to be present are diregarded. Thus, whether analyzable expression of this kind are typically formed with the dependent constituent in the genitive case, as in Kildin Saami, whether there is a free-standing word expressing the association of the two constituents as in example (6g.) from Swahili, or whether the typical structural template for the formation of complex expressions involving two stems requires the presence of such a marker *and* reduplication of the head, as in (6h.) from Fijian, or whether any other grammatical material or process next to two lexical stems is present, is disregarded: what matters for typologization is the presence of two elements with lexical meaning that have the potential to occur freely (though note that North American-type lexical affixes are treated along with these). Note also that this entails that additionally present derivational markers do not affect classification of a given form as being of the lexical type.

While the manner of classifying complex forms proposed here does not require a clear-cut distinction between compounds and lexicalized phrases, it does require some knowledge about the potential of the involved morphemes to occur as free-standing forms to distinguish between derived and complex lexical expressions, because this distinction is based on the number of lexical morphemes that are present. One problem for this classification, then, is the very common grammaticalization path by which certain lexemes which are used frequently in compounds or phrases come to lose their semantic content over time and develop into derivational morphemes. For instance, the present day German derivational morpheme *-heit*, used to derive abstract nouns from adjectives and nouns, originates in the full Proto-Germanic noun **haidu* 'manner, appearance' (Kluge 2002).[18] Indeed, there are several languages in the sample which have a class of morphologically complex terms formed according to a quite frequent recurrent pattern involving a certain word, as in (7.) and (8.):

[18] 'Art und Weise, Erscheinung' is the original German gloss.

(7.) Toaripi lexemes with *fare* 'fruit'
a. *ori fare* 'bird fruit' = 'egg'
b. *kōu uti fare* 'back bone fruit' = 'kidney'
c. *kō fare* 'scrotum fruit' = 'testicle'

(8.) Chickasaw lexemes with *oshi'* 'son'
a. *akankoshi'* /akanka'-oshi'/ 'chicken-son' ='egg'
b. *okkisoshi'* /okkisa'-oshi'/ 'door-son' = 'window'
c. *ilbak-oshi'* 'arm/hand-son' = 'finger'

In Toaripi, complex terms with *fare* 'fruit' are commonly used to denote objects that are roundish in shape, and in Chickasaw *oshi'* 'son' occurs frequently in terms for comparably small things. It is therefore natural to wonder whether Toaripi *fare* and Chickasaw *oshi'* can be considered to have their full lexical meaning in the complex terms in question or whether they are rather used (or are on their way to being used) in a derivational fashion, with the semantics bleached to something like 'roundish thing' and 'small thing' respectively (indeed, it is common cross-linguistically for diminutives to develop from words meaning 'child,' Jurafsky 1996). This is a potential problem for classification, but one for which provisional criteria are available to make analytic decisions in a principled way: grammaticalization often (but not always, at least not in the initial stages of the process) involves phonological reduction. Therefore, if the questionable morpheme is not different in its phonological structure from a free-standing form with lexical meaning in the consulted source, the form was analyzed as being of the lexical, not the derived type. Second, the very fact that there is a lexeme with lexical meaning listed in the source provides evidence that indeed this lexical meaning is available to speakers and (at least) potentially perspicuous to speakers when occurring in the complex structure. Both is the case for Toaripi *fare* and Chickasaw *oshi'*. Consequently, the examples in (7.) and (8.) were analyzed as lexical, not derived.

A further type of formal relation that is of less importance cross-linguistically, but still needs to be distinguished as a separate type is SYSTEMATIC ALTERNATION OF GRAMMATICAL PROPERTIES which distinguishes different meanings of essentially the same phonological form. Examples are in (9.).[19]

[19] Another possible class of lexical relation that is theoretically included in this rubric is systematic tone alternation. There are some potential examples that may be instances of this, compare for instance Sko *hì* 'blood' and *hí* 'resin.' However, this may also be due to sheer accident, since there are no other examples of the same technique to

(9.) a. Number Alternation
Bezhta *häy* = 'eye;' *häydä* 'eye.PL' = 'glasses'

b. Gender Alternation
Embera *kidhátrɨ* 'chin' (masc.); 'jawbone, mandible' (fem.)
Khoekhoe *ǂare-s* 'tail/buttocks-3SG.MASC' = 'tail'
ǂare-b 'tail/buttocks-3SG.FEM' = 'buttocks'

c. Noun Class Alternation
Kiowa *tɛ̀i-p* 'calf/buttocks-NOUN.POSTFIX' = 'calf'
tɛ̀i-dl 'calf/buttocks-NOUN.POSTFIX' = 'buttocks'

Lumping these different processes together into one single category is questionable given the diverging nature of each single subtype; in spite of this, it is done because of the relative paucity of examples of each type in the database. The cover term ALTERNATING will be used to describe each of the subtypes.

Finally, with regard to terms such as (6h.), from Rama, it will be said that they are ANALYZABLE, but semantically REDUNDANT, and therefore not motivated according to the definition: *ngárkali ~ ngarkalima* alone already means 'flame' and *abúng* 'fire' does not add any further apparent semantic content (note that redundancy in this sense is a purely formal notion, and does not necessarily rule out that the complex term fulfils pragmatic or other functions).

3.6.2. SEMANTIC CLASSIFICATION

This section develops the approach to the classification of semantic relations adopted in this study. In doing so, the discussion attempts to weave together various threads from different approaches to semantic relations. These include traditional accounts of semantic relationships of noun-noun compounds, Cognitive Linguistics, and Cognitive Psychology. Last but not least, reference will be made to the grid developed by Koch (2001) and Koch and Marzo (2007) to arrive at a cross-linguistically workable inventory of semantic relations. It should be noted right at the outset that cognitive or psychological reality of the semantic analyses made is explicitly not claimed, in spite of adducing literature from Cognitive Linguistics and Cognitive Psychology. Arguably, waiving claims as to psychologi-

---

connect lexical meanings by tone alternation, and such isolated instances were therefore disregarded in the light of absence of evidence for systematicity of the attested cases.

cal reality does not absolve the analyst from providing a framework that is in line with findings of related disciplines and to allow him/her to set up analytic categories as s/he pleases.

3.6.2.1. Review of previous accounts

3.6.2.1.1. *Studies of semantic relations in Noun-Noun Compounds.* The discussion departs from proposed classifications of semantic relations in noun-noun compounds in individual languages. Quite a large number of different accounts of semantic relations within these have been proposed. Table 1 provides an overview:

	Proposed Relations
Li and Thompson (1981); wording at times altered	place of location, place of application, use, unit, piece of equipment used in sport, protective device against, causation, container, parallel, product, material, place where sold, disease of, time for, source of energy, metaphorical description of, component of, source of, employee or officer of, proper name, person who sells or delivers
Marchand (1969)	comparison, material a thing is made of, purpose, place, time
Downing (1977)	whole-part, half-half, part-whole, composition, comparison, time, place, source, product, user, purpose, occupation
Levi (1978)	CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM, ABOUT
Warren (1978)	source-result, part-whole, origin in time or space, purpose, activity-actor, resemblance

table 1: suggested inventories to describe semantic relations between the members of Noun-Noun Compounds

As becomes immediately obvious, the different inventories of semantic relations vary with respect to their granularity: Li and Thompson's (1981) inventory of proposed relations within Mandarin Chinese noun-noun compounds is larger and contains more specific relations than does, say, Marchand's (1969). Also, the classifications vary with respect to their purpose: While Li and Thompson (1981: 49) find it simply "pedagogically and heuristically useful to describe them," Levi (1978) is a study in the framework of Generative Semantics that aims to show that compounds can be derived from underlying propositions (sentences) by means of deletion of an underlying predicate, whose semantics is said to be captured by the relations she posits.

The classifications also vary with regard to their scope. Levi (1978), for instance, restricts the scope of her study to endocentric compounds from the start and therefore her inventory lacks the 'metaphorical description/comparison/resemblance'-category used to describe relations within exocentric compounds in the other classifications. To be fair, with the exception of Levi (1978: 75), who claims that "the variety of these relationships is in

fact confined within a very limited range of possibilities," all above quoted authors make explicit that their proposals are not meant to be exhaustive and do not aim at appropriately describing all semantic relations between two members of a compound that could ever be found. In fact, Marchand (1969: 22) states that "[i]t is no use trying to exhaust the possibilities of relationship" and that "many compounds defy an indisputable analysis," echoing in a moderate version the (in-)famous statement made by Jespersen (1942: 143) that "the number of possible logical relations between the two elements is endless." Indeed, it can be argued that this is the case, given that the semantic relation within two given compounds will always be subtly different every time, and Plag (2003: 148) notes that "such semantically based taxonomies appear somewhat futile" because of their inherent arbitrariness. What is desired, however, is a useful way to abstract from individual cases, and this certainly was the goal of the above mentioned authors, as is the goal within the present study.

Unfortunately, the proposed inventories are not directly applicable to the task of the present study, for a variety of reasons. First, in some cases, the semantic relations are sometimes very specific (for which researchers can hardly be blamed, as they were trying to account for recurrent patterns in the languages they were describing). More importantly, however, the suggested categories do not describe the present data exhaustively, and it is desirable to find a semantic grid which is general enough to describe all semantic relations that surface in the database on which this study is based. For instance, it seems difficult to assign Rendille *daáyto* 'pupil,' a compound consisting of *daáyi* 'black' and *to* 'thing,' to any of the categories in any of the above classifications. What is more, only compounds are considered in the reviewed proposals, whereas for the present task also semantic relations in derived and colexifying lexical items are to be described. However, it is worth bearing in mind for the subsequent discussion that mostly a basic distinction in the classification made by earlier authors can be spotted: on the one hand, there is the relation of 'metaphorical description of' (Li and Thompson 1981), 'comparison' (Marchand 1969, Downing 1977) and 'resemblance' (Warren 1978) to describe semantic relations within exocentric compounds. This relation is not very elaborate in all schemes. Instead, the above mentioned studies at large concentrate on semantically "harmless" endocentric compounds at the expense of a detailed analysis of the metaphorically driven ones which are presumably more difficult to describe. On the other hand, all other proposed criteria can be grouped into one large class, in which the relation between the elements of the compounds, typical for the endocentric type, are not based on some perceived similarity or resemblance but rather on some more immediate semantic relation.

3.6.2.1.2. *Qualia Structure.* Qualia structure is one of the three dimensions argued for in Pustejovsky (1995) for the semantic analysis of lexical items of a given semantic category within the more general theory of the "Generative Lexicon," an approach to lexical semantics in principle compatible with computational implementation. It deserves discussion not only because the notion of qualia structure is outlined mostly with reference to nominals in Pustejovsky (1995), but also because it is a semantic account that allows to elegantly deal with lexical items related by word-formation. One of the advantages of introducing qualia structure is that it allows "nouns, and consequently the NPs containing them, to encode information about particular properties and activities associated with the them [sic!]" (Pustejovsky 1995: 79). This is one of the major novelties of the Generative Lexicon approach, because qualia structure thus provides the verbs with information that is required to arrive at a consistent semantic interpretation of the clause as a whole, and thus distributes the information for semantic interpretation of clauses more evenly among its elements than other heavily verb-based approaches to clausal semantics. There are four qualia structures in Pustejovsky's approach, though not all need always be specified. These are CONSTITUTIVE, FORMAL, TELIC, and AGENTIVE (Pustejovsky 1995: 85-86). The constitutive quale specifies the relation between an object and its constituents, such as material, weight, and parts and component elements, whereas the formal quale is "[t]hat which distinguishes the object within a larger domain." Examples mentioned by Pustejovsky are dimensions such as orientation, magnitude, shape, dimensionality, color, and position. The telic quale is easiest to summarize: it specifies the purpose and function of an object, while, complementary to the telic quale, the agentive quale spells out "[f]actors involved in the origin or 'bringing about' of an object." Thus, a *novel* is said to be characterized by the agentive quale writing and the telic quale reading. Pustejovsky (1995: 89) is now in a position to describe the semantics of derived nouns such as *typist* by simply stating that "the TELIC makes direct reference to the process-denoting verb from which its nominal is derived." Similarly, for compounds, the telic role in nouns such as *book shelf* can be viewed within the theory of qualia structure to be specified by the dependent element *book*. Thus, although conceived of from a quite different and more formal angle, Pustejovsky here approaches traditional accounts to compound semantics as described above.

3.6.2.1.3. *Cognitive Psychology*. This section briefly reviews relevant literature for the topic at hand from Cognitive Psychology, beginning with psychological evidence for the notion of semantic relatedness as evidenced by semantic networks revealed by lexical priming (a strand of research initiated by Collins and Quillian 1969) and moving on later to a discussion of conceptual combination, which is a topic related to morphologically complex expressions.

Collins and Quillian (1969) have shown that encyclopedic information about concepts that are stored together with a primed concept can be retrieved faster the closer the association is. Moss et al. (1995) conducted priming experiments for three types of lexical activation, namely co-taxonomic or taxonomic relations, functional relations (e.g. that between 'broom' and 'floor') and what they call "script-related" relations (i.e. what others would call relations of contiguity, e.g. the relation between 'restaurant' and 'wine'). They found that the latter type of relation yielded only marginally significant results in priming experiments, concluding that this is a less central aspect of semantics than other types of information. However, even in the psychological literature, what the category of "associatively related items" should be taken to include is often not deduced from experimental data. Instead, the knowledge of what is and what is not associatively related is itself axiomatically presupposed as a prerequisite to the experiments. Revealing in this context is that Neely (1991: 294) speaks of nodes for both semantically and associatively related words being " 'stored close together' " (single quotes in the original), which implies that it is not at all clear what the precise psychological reality of being "stored close together" is. Furthermore, psycholinguistic studies of word sense discrimination appear to make use of the microstructures of dictionary entries to a surprising degree. According to Miller et al. (1988: 4), as quoted in Jorgensen (1990: 168), "[b]y and large, psycholinguistic experiments presuppose the validity of the general structures that linguists and lexicographers have identified and try instead to test hypotheses concerning the way such structures arise or how they contribute to other cognitive processes." This, of course, limits the exploitability of such psycholinguistic approaches for the present task, because their experiments are not independent of potentially arbitrary and unprincipled decisions made by lexicographers in the process of dictionary making.

Next to these approaches, there are also accounts of adjective-noun combinations that operate within prototype theory as initiated by Rosch's (1978) research. Smith et al. (1988) discuss the possibility to decompose prototypes in attribute-value pairs. An attribute is "a concept that describes an aspect of at least some category members. For example, *color* describes an aspect of birds, and *location* describes an aspect of *vacations*" (Barsalou 1992: 30). Smith et al. (1988) propose that a modifying adjective in an NP can be seen as overriding the inherent prototype value for the respective attribute relevant for the head noun. For instance, 'apples' can be argued to have prototypically a red color, which would be the default value inherited from the prototype without further adjectival modification. In the phrase *green apples*, then, the adjective *green* selects the relevant attribute in the noun (color in this case) and overwrites the prototypical value.

Another research current of Cognitive Linguistics that is relevant for the present context are studies concerning so-called "conceptual combination," because they are

intricately connected with morphologically complex expressions. The basic question researchers engaging in this field ask is according to what psychological rules people combine existing concepts (expressed by lexical items) to coin neologisms and what procedures they apply in deducing the semantics of a novel word from its constituent parts. There are at least three concurring theories with regard to this question, the dual process theory, as forwarded e.g. by Wisniewski (1996, 1997), the relational theory (e.g. Gagné and Shoben 1997), and the constraint theory (e.g. Costello and Keane 2000, 2001). Brief discussion of all these theories, which however cannot do full justice to details, follows (see also Estes 2003: 305-308, who influenced this discussion).

According to dual process theory, novel attributive compounds are understood by means of a sequence of two distinct cognitive mechanisms of comparison and following attribution. Thus, in hearing the noun-noun compound *cactus carpet*, the concepts 'carpet' and 'cactus' are compared with each other. This comparison then determines which property will be mapped and where it will be mapped to. The comparison will reveal that carpets are typically soft while cacti are typically prickly, which is one of the key features in which the two concepts differ. The second step of the comprehension and interpretation process that follows the comparison is the attribution, which is why such interpretations are called attributive interpretation; in this particular case the "prickliness" of cacti will be mapped as an attribute of the carpet as a feature that sets it off from other carpets (example from Estes 2003). In the dual process theory, so-called "relational combinations" such as *floor television* are said to be interpreted by a completely distinct mechanism, namely assignment of theta roles. In the above example, *floor* would be assigned the theta role of location, and the head noun television is that which is located. To make the distinction between attributive and relational interpretations clearer, another example may serve: in Wisniewski and Love's (1998) study, some participants interpreted *book magazine* as 'a magazine about books.' This is a relational interpretation. Others, however, interpreted this compound as meaning 'a thick magazine,' which is an attributive interpretation, in which a property of books, namely being thick, is transferred to the magazine in question.

The competing relational theory does not posit two distinct processes for the comprehension of conceptual combination. Instead, it is argued, "nominal combinations are understood by inferring some relation that is purported to exist between the constituent concepts" (Estes 2003: 306). This approach rejects the distinction between attributive and relational processes, and instead views attributive relations, which it does recognize, as a particular relational process, namely that of resemblance.

A third approach - constraint theory - posits, as its name suggests, three constraints on the interpretation of novel compounds, namely, the plausibility constraint, the

informativeness constraint, and the diagnosticity constraint, which "requires the construction of an interpretation containing diagnostic properties from each of the concepts being combined," and which is the most important constraint in this framework (Costello and Keane 2001: 257). Thus, the compound *cactus fish* is more likely to be interpreted as meaning 'a prickly fish' rather than 'a green fish,' since "*prickly* is more diagnostic of cactus than *green*." (Costello and Keane 2001: 257).[20] Given that each novel compound can be assigned a very large variety of interpretations, the constraints are a modeling of how the best interpretation is arrived at by the hearer.

Each theory makes different predictions as to how novel compounds will be interpreted, which are not of further concern here. More important for the present purposes is that relational interpretations (as when the compound *apartment dog* is interpreted as meaning 'a small dog that lives in city apartments') and property interpretations (as when *elephant fish* is interpreted as meaning 'a big fish') bear a striking resemblance to the notions of similarity and contiguity, the mechanisms commonly assumed to be underlying the phenomena of metaphor and metonymy (though Cognitive Psychologists might disagree). When *elephant fish* is interpreted as 'big fish,' this fish is perceived to be similar to an elephant in some respect. When *apartment dog* is interpreted as 'small dog living in apartments,' the typical location such dogs are encountered in are apartments; thus, the meaning of *dog* and that of *apartment* are in a relationship of spatial contiguity.

However, it should be borne in mind that the cited studies all deal with fabricated novel compounds, not institutionalized lexemes of the participants' language (English). Murphy (1988: 530) is very clear about this: perhaps surprisingly, he makes an a priori equation of morphological simplicity with conceptual simplicity by defining that a "concept is 'simple' if it can be represented as a single lexical item" and "a concept that requires more than one lexeme is 'complex'." However, he hastens to add to the second definition "[u]nless its linguistic expression is lexicalized (i.e., idiomatic)" and explains that this qualifier "serves to rule out idiomatic phrases like *dog house*, which has the conventional meaning 'house that a dog sleeps in,' and therefore may no longer be a truly complex concept." This limitation is probably to some extent borne out of the psychologists' need for controlled experimental settings in which participants of studies are unbiased in that they are confronted with complex terms they have not heard before. It is obviously most unfortunate for the context of the present study, because it deals precisely with such conventional expressions, not neologisms.

[20] The reason why different studies are equally fond of using properties of cacti as examples is not clear.

3.6.2.1.4. *Diachronic Semantics.* Possible avenues towards a classification of semantic relations are not only suggested in synchronic studies from different theoretical perspectives, but also in studies of semantic change. Matisoff (1978) employs a cross-classification of semantic associations in diachrony according to whether the change involves a shift in the semantic domain (trans-domain change) or not (intra-domain change). An example of the former would be a semantic shift from 'stomach' (in the semantic domain of body-parts) to 'cave' (in the semantic domain of topological concepts). An example of the latter would be when a body-part term shifts in meaning to denote another body-part term, with the semantic domain remaining the same. Further, Matisoff distinguishes between, roughly put, metonymic and metaphoric changes, with further subdivisions that are particularly tailored for application to the domain of body-part terms. Interestingly, Matisoff, leaving the primarily diachronic orientation of his classification somewhat, also discusses "association via compounding." This illustrates nicely the sometimes intricate commonalities as far as semantics is concerned that can be observed in colexifying and complex lexical items noted in § 3.5.[21] Matisoff's approach is adopted, in a terminologically modified version, in Wilkins (1996), and is applied there exclusively to semantic change. As Riemer (2010: 376) summarizes, "[t]he centrality of metaphor and metonymy in semantic change is due to the fact that they jointly exhaust the possibilities of innovative word use and thus subsume all the other descriptive categories." Notably, however, semantic change does not come about suddenly, but rather involves an intermediate stage in which both the original and the innovative sense are available for the same linguistic sign (Wilkins 1996, Traugott and Dasher 2002, Evans 2010). The lexical semantics of polysemy/colexification in synchrony, in turn, is central to the present study, and there is thus no principled reason why it should not be possible to accommodate frameworks for studying diachronic semantic change to synchronic questions and vice versa.

3.6.2.1.5. *Meaning-text theory and semantic derivation.* Another relevant line of research is that of the meaning-text theory associated most prominently with the name of Igor A. Mel'čuk. Meaning-text theory is a formal framework for the description of natural language in which semantic representations (meanings) are mapped onto final phonological representations (texts) via a number of intermediate stages (an outline is in Mel'čuk 1981). One of the points that sets meaning-text theory apart from other such frameworks is that it assigns a central role to the lexicon. Mel'cuk and Polguère (1987: 265) turn the

[21] Similarly, François (2008) mentions the possibility that his semantic maps showing exclusively associations by colexification be amended by associations by word-formation.

Bloomfieldian slogan of the lexicon as an appendix to grammar around by saying that in meaning-text theory the grammar, conversely, is an appendix to the lexicon. Hence, work in terms of meaning-text theory incorporates a detailed description of semantics of lexical items and their combinatorial possibilities. This rich information is collected in so-called Explanatory Combinatorial Dictionaries, compiled by Mel'čuk and associates for Russian and French. A crucial role, in particular in the modelling of cooccurence restrictions, is played by so-called lexical functions. These include traditionally recognized lexical relations; for instance, the lexical function called "Syn" returns, if the argument *to shoot* is passed to it, *to fire*. There are all in all more than 50 such lexical functions, and some relate to another (relatively marginal) concept within meaning text-theory, namely semantic derivation. This is conceived of as a purely semantic counterpart to morphological derivation. Thus, as Mel'čuk's (2007: 120-121) example runs, Russian *stolovaja* 'dining room' is derived morphologically from *stol* 'table,' but semantically from *est'* 'eat.'[22] A prerequisite for semantic derivation in this sense to be diagnosed is that the semantic association is regular and found in other like lexical pairs within the same language.

But the notion of semantic derivation is also applied more broadly; in fact, in Apresjan (1992: 194) semantic derivation appears as a synonym for polysemy (see Apresjan 1992: 209-211 for references to literature which brought about this broadening). The term also plays a role in cross-linguistic approaches to the lexicon in the Russian linguistic tradition. For instance, in Zalizniak et al. (2012) the term is used akin to polysemy, but in a slightly broader way, namely to describe the relationship between two related meanings of a word that can be considered as derivative from one another, on the one hand in the form of general semantic templates of polysemy present in a given language, but also in the form of non-systematic semantic associations as long as they recur with some frequency cross-linguistically. In this use, semantic derivation also includes diachronic aspects (cf. the term semantic shift in Zalizniak 2008, Zalizniak et al. 2012, which is also construed broadly to include semantic association by way of polysemy, word-formation, diachronic semantic change, and some other types). In the first mentioned sense, the term is identical to Apresjan's (1974) "regular polysemy," called "regular" because part of the definition is that

[22] Mel'čuk (1976c) in a similar vein argues that formal complexity in derivation need not always correspond to semantic complexity. Hence, *katit'sja* 'to roll oneself' is derived formally from *katit'* 'roll,' but semantically *katit'* is said to be more complex in that it is inherently causative. This Mel'čuk (1976c: 70) calls "inverse word-formation" ("inverse Wortbildung;" "obratnoe slovoobrazovanie"). There is also the related notion of derivational suppletivism ("derivationeller suppletivismus") mentioned by Mel'čuk (1976c: 68): *vrač* 'doctor' is in the same relation to *lečit'* 'heal' as *spasitel'* 'saviour' is to *spasti* 'save.' Mel'čuk (1976c: 78) suggests to always differentiate between morphological and semantic facts in derivation (see also Mel'čuk 1976b, Apresjan 1992: 184-185).

the general abstract relationship between polysemous readings occurs in more than just one example in a language (as with Mel'čuk's semantic derivation). Apresjan (1974: 17-18) also explicitly recognizes the close relation between (regular) polysemy and word-formation along a variety of parameters that is relevant for present purposes as seen in § 3.5. (compare also Apresjan 1992: 209-211 for different views of authors in the Russian tradition concerning the relationship between polysemy and word-formation). Apresjan (1974) importantly also operates with the distinction between metonymy and metaphor, stating that in regular polysemy meanings stand most often in a metonymic relationship to one another, while in irregular polysemy metaphor is more frequent. Regular polysemy in this sense is also what Mel'čuk (1973: 111) has in mind, when talking about "polysemy of the type of Russ. *sliva* – 'plum' – 'plum tree' (*gruša* 'pear' – 'pear tree', *višnja* 'cherry' – 'cherry tree') and in general, the widespread phenomenon of the same stem having both the meaning X and 'something connected in a definite way with X', for instance an action and its result (*nagnoenie* 'festering'), an action and its place (*mojka* – 'washing' and 'washing place'), an action and its object (*vyšivanie* 'embroidering' and 'embroidery'; *trebovanie* 'requiring' and 'requirement'), an action and its instrument (Engl. *intake*), etc." He even suggests use of the term "semantic derivation" for such regular polysemies, but since this term is already used in meaning-text theory, suggests the label "semantic conversion" instead. [23]

3.6.2.1.6. *Cognitive Linguistics.* The rise of Cognitive Linguistics in the early 1980's, with works such as Lakoff and Johnson (1980) and Langacker (1987a), made available a fundamentally new perspective on semantic phenomena (which seem at least partly reflected in Wilkins's 1996 typology of semantic change). Particularly influential was Lakoff and Johnson's (1980) notion of conceptual metaphor, which, as they have argued in detail, pervades thinking and links whole areas of thought with each other. Metaphor, according to a standard Cognitive Linguistics definition, can be understood as a "cognitive mechanism whereby one experiential domain ... is partially mapped onto a different experiential domain, the second domain being partially understood in terms of the first one. The domain that is mapped is called the *source* or *donor domain*, and the domain onto

[23] This discussion is embedded in the establishment of a general typology of form-meaning pairings. Mel'čuk (1973) states that the relationship between two forms or two meanings can only be of four kinds: identity, inclusion, intersection (existence of common part), and absence of common part, which yields a cross-classification of different types of formal and semantic relations with 17 distinct scenarios, some of which model classical categories such as homonymy, others particular types of derivation or even more special cases, such as the aformentioned semantic conversion.

which it is mapped, is called the *target* or *recipient domain*" (Barcelona 2002: 211). A textbook example of metaphor is the ARGUMENT IS WAR metaphor discussed by Lakoff and Johnson (1980: 4), which accounts for the appropriateness of sentences like *Your claims are indefensible, he attacked every weak point in my argument, his criticisms were right on target, I demolished his argument,* and so on: expression from the domain of arguing (*claims, criticisms, arguments*, etc.) are mapped to the domain of war (*indefensible, attack, demolish*, etc.). Note that metaphor is explicitly not a linguistic, but a cognitive phenomenon in this view: linguistic utterances merely reflect, but do not constitute, metaphors such as ARGUMENT IS WAR.

Metonymy was initially the step child of Cognitive Linguistics, but has more recently (e.g. Kövecses and Radden 1998) been at the focus of attention within the research paradigm. It has come to be conceived of as "one of the most fundamental processes of meaning extension, more basic, perhaps, even than metaphor" (Taylor 2003: 126), as several authors have suggested that many metaphors have a metonymic basis. Metonymy is formally understood to be a mechanism "whereby one experiential domain is partially understood in terms of another experiential domain included *in the same common experiential domain*" (Barcelona 2002: 215). Examples (except for the last, which is common lore, from Kövecses and Radden 1998) include *the buses are on strike* (in which *buses* refers not to the buses themselves, but to the busdrivers operating them), *the White House did not intervene* (in which *the White House* is used metonymically to refer to the government it hosts), and the infamous *the ham sandwich wants the bill*, in which *ham sandwich* is meant to refer to the restaurant guest who has previously ordered it

However, despite the new possibilities it opens up for semantic description, Cognitive Linguistics has relatively little to say about the nature of semantic relations in lexical items, i.e. conventionalized expressions as opposed to ad-hoc processes. There are, however, some notable exceptions. As Dirven (1985: 87-88) writes, "[o]ne of the many self-imposed limitations found in linguistic and philosophical writings is that metaphor is only seen at the sentential level ... If one starts from a view on metaphor as a ubiquitous cognitive process, one would rather expect metaphor (in its broader sense) to be operative at various levels of language structure and linguistic units, and not just at that of the sentence." In a sense, the focusing of Cognitive Linguistics on sentential processes is of little surprise: Cognitive Linguistics is mainly concerned with the dynamic mental construal of extra-linguistic situations by cognition and its mirroring in linguistic structure. The non-dynamic nature of lexical items and the relatively fixed semantic relations that may hold between members of complex lexical items or different senses of lexically entrenched polysemy therefore seem to be simply of little interest to the

Cognitive Linguistic enterprise. This position is present from the very beginning of the development of the field. Lakoff and Johnson (1980: 54-55) say:

> In addition to these cases, which are parts of whole metaphorical systems, there are idiosyncratic metaphorical expressions that stand alone and are not used systematically in our language or thought. These are well-known expressions like the *foot* of the mountain, *a head* of cabbage, the *leg* of a table, etc. These expressions are isolated instances of metaphorical concepts, where there is only one instance of a used part (or maybe two or three). Thus the *foot* of the mountain is the only used part of the metaphor A MOUNTAIN IS A PERSON. ... The point here is that there are metaphors, like A MOUNTAIN IS A PERSON, that are marginal in our culture and our language; their used part may consist of only one conventionally fixed expression of the language, and they do not systematically interact with other metaphorical concepts because so little of them is used. ... They ... are not metaphors that we live by.

Of equally little surprise, then, is that cognitive approaches to word-formation are not so much concerned with actually existing word-formation items and their morphological and semantic structure, but more with abstract schemas available to speakers which allow them to coin new compounds (Tuggy 1987, 2005) or with an, itself metaphorical, extension of the well know figure-ground organization to the domain of word-formation (Ungerer 2007). Ungerer (2007: 671), in fact, states that "current cognitive research in word-formation is still very much in its initial stages" and that "the application of most empirical methods has been too selective for a proper evaluation of their usefulness."

Still, it is not at all the case that Cognitive Linguistics analyses cannot be fruitfully exploited in the present context. Cognitive Linguistics has, by redefinition of long-standing classical ideas, revived the powerful analytic notions of metaphor and metonymy, and in fact, these have been applied to account for ad-hoc semantic extensions of particular lexemes at the utterance level and sometimes also to certain cases of conventionalized lexico-semantic extensions (e.g. Lakoff 1987, see also Svanlund 2007). Both notions are widely used and have come to be the quasi-standard means of describing semantic extensions (though see Levinson 1994 for critique of analyses of extensions of body-part terms in terms of metaphor). Now, analyses on the utterance level are not in principle different phenomena from lexically entrenched meaning extensions, because the latter typically arise out of the conventionalization of the former. Cognitive Linguistic analyses, however, including those involving metaphor and metonymy as descriptive categories, have been criticized for their speculative nature (Riemer 2010: 255). In addition, the notions of metaphor and metonymy would need to be accommodated to be applicable to the slightly shifted context of the present study, with this accommodation ideally at the same time eliminating some of the indeterminacy of traditional Cognitive Linguistics. An attempt to achieve these goals is presented in the following section.

3.6.2.2. Metaphor and Metonymy within the lexicon

The analyses to be made presently pertain to the lexicon, conceived of as an inventory of fixed conventionalized expressions. Therefore the notions of metaphor and metonymy need to be slightly altered to be applied to lexical items, and to ultimately make available similar tests as commonly employed in lexical semantics to diagnose phenomena in individual languages (see e.g. Cruse 1986). Test frames would also reduce a felt arbitrariness in analytical decisions to a significant degree.

One proposal that goes into this direction in fact exists: in order to test for the presence of metaphor at the utterance level, Ray Gibbs, as quoted in Kövecses (2002: 146), suggests the so-called 'is like'-test to identify similarity-based conceptual mappings.[24] Examples are:

(10.) a. The *creampuff* was knocked out in the first round of the fight.
b. We need a new *glove* to play third base.

In (10a.), *creampuff* refers to a boxer in a boxfight, while in (12b.), *glove* refers to a baseball player. Gibbs's 'is-like' test makes explicit the underlying semantic processes of those substitutions:

(11.) a. The boxer is like a creampuff.
b. *The third baseman is like a glove.

(11a.) is acceptable, and thereby (11a.) is diagnosed as a metaphorical substitution, while (11b.) is not a felicitous statement in English, and is therefore not a similarity-based substitution (it is said to be automatically metonymic in Kövecses 2002; as will be discussed below, in the present framework a separate test will be used).

This test is well suited not only for utterance-level metaphorical and metonymical substitutions, but is also applicable to semantic relations in colexifying and complex expression, although it was not originally designed for this purpose. It will be used, with some modifications that formalize the application of the test somewhat, as the prime

[24] There appears to be little agreement between scholars in Cognitive Linguistics as to the status and necessity of the notion of similarity in metaphor (see e.g. Grady 1999 for an overview). Note that the very notion of metaphor being based on similarity is explicitly rejected by Lakoff and Johnson (1980), although also in recent work, such as Riemer (2005), contiguity and similarity are at the heart of Cognitive Linguistics analyses. For a defence of similarity, see also Murphy (1996).

device to identify metaphor-based relations between different senses of colexifying lexical items and constituents of complex expressions here. Consider the following colexifying words one of whose dictionary glosses is 'beard':

(12.) a. Arabela *mohua* 'beard, antenna of insect'
b. Tetun *timir* 'beard, chin'

Applying the 'is like'-test to different senses of a colexifying lexical item involves inserting the glosses into the general template of the test frame:

(13.) <gloss 1> is like <gloss 2>

Thus for Arabela *mohua* in (12a.) one gets 'a beard is like the antenna of an insect,' which is a semantically acceptable statement drawing attention to certain similarities between beards and antennae (they are both structures that protrude from the head of a living being, etc.). In contrast, for (12b.), one gets 'a beard is like a chin.' Clearly, this is an odd and infelicitous statement, as the relation between 'beard' and 'chin' is not one of being like each other, but rather characterized by the fact that beards grow on chins. The 'is like'-test has thus shown that (12a.) can be characterized by a similarity-based metaphor, and has demonstrated that (12b.) is not so describable. One can also use the 'is like'-test for morphologically complex expressions. Consider the examples in (14.).

(14.) a. Cashinahua *kex-ni* 'mouth-forest'[25]
b. Dadibi *penani nisi* 'chin hair'

Using the 'is like'-test on the level of lexical analysis involves inserting the overall meaning of the complex term and each of the constituents of the complex term with lexical meaning into the general template of the test:

(15.) <meaning of term> is like <constituent$_{1,2,...,n}$>

Doing so for each of the four constituents in the examples in (14.), this only yields a fortuitous result in one case, namely with the element meaning 'forest' in Cashinahua: for (14a.), one gets 'a beard is like a forest' and 'a beard is like a mouth,' and for (14b.) 'a beard

[25] In fact, *kex-* is a prefix that can also refer to the 'lips' as well as more abstractly to an 'edge' or 'contour.'

is like hair' and 'a beard is like a chin.' (14a.) is a felicitous statement about a perceived similarity between some properties of forests with some properties of beards (they are both dense agglomerations without any inherent structure of individual entities, hairs and trees), and by virtue of this, the 'is like'-test yields a positive result (it is enough for metaphor to be diagnosed if this is the case with one of the constituents). The other statements, in contrast, are odd semantically: a beard is not like hair, if anything it is a kind or a certain configuration of hair, and so on. The results of the 'is-like' test appear to be reconcilable with Gentner's (1983) account of metaphor, where metaphor is thought of as a special subtype of analogy as a more general device of cognitive structure mapping.

The 'is-like'-test allows describing the semantic relations in (12a.) and (14a.), but what about (12b.) and (14b.)? These cases are instances of what would intuitively be called contiguity-based semantic relations: 'beards' are spatially contiguous to 'chins' and to 'mouths' in that this is the place where they grow, and they are a particular configuration of hair. However, the notion of contiguity is, while intuitively appealing and looking back to a millennia-long tradition,[26] oddly ill-defined (Geeraerts 2010: 27), and this vagueness in definition remains to some extent to the present day (Seto 2003). Koch and Marzo (2007: 262), for instance, merely state that it is "the fundamental connection underlying all kinds of frames, scenarios, scripts etc. and including part-whole relations," a definition which recurs to the equally vague notions of frames, scenarios and scripts, however widespread and useful they may be.[27]

In the present context, therefore, a complementary test to the 'is like'-test will be used in order to diagnose contiguity-based relations that is foreshadowed in Mel'čuk (1973: 111). This is the 'has something to do'-test, which works analogously to the 'is like'-test in

[26] Contiguity as a technical term goes back to Aristotle's treaty on remembering Περὶ μνήμης καὶ ἀναμνήσεως, also known as de memoria et reminiscentia, where the three mnemonic principles contiguity, similarity, and contrast are established ("ὅταν οὖν ἀναμιμνησκώμεθα, κινούμεθα τῶν προτέρων τινὰ κινήσεων, ἕως ἂν κινηθῶμεν μεθ' ἣν ἐκείνη εἴωθεν. διὸ καὶ τὸ ἐφεξῆς θηρεύομεν νοήσαντες ἀπὸ τοῦ νῦν ἢ ἄλλου τινός, καὶ ἀφ' ὁμοίου ἢ ἐναντίου ἢ τοῦ σύνεγγυς"/ "Whenever, therefore, we are recollecting, we are experiencing certain of the antecedent movements until finally we experience the one after which customarily comes that which we seek. This explains why we hunt up the series (of *kineseis*) having started in thought either from a present intuition or some other, and from something either similar, or contrary, to what we seek, or else from that which is contiguous with it," Beare's 1908 translation). These principles were elevated to the status of "laws of association" in the philosophy of the British empiricists such as Hume and Locke, and from there have made their way to the slowly emerging psychology as an independent field of research (see Warren 1921 for an excellent overview) and to linguistics thanks to Jakobson (1956/1971).

[27] Compare e.g. Fillmore (1982: 111): "By the term 'frame' I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits."

analyzing colexifying lexical items, with the difference that the additional statement "but is/are not similar to it/them" is added to force the contiguity-reading "of having something to do" (after all, 'beards' also "have something to do" in one sense with 'forests' in that a certain similarity between them may be perceived).

(16.) <gloss1> has something to do with <gloss2>,
but is/are not similar to it/them.

Thus for (12b.), one gets 'beards have got something to do with chins, but are not similar to them' which is a semantically acceptable statement. For complex terms such as (14b.), the 'has got something to do'-test is applied by inserting the overall meaning of the expression and again each of the constituents into the test frame.

(17.) <meaning> has something to do with <constituent$_{1,2,...,n}$>,
but is/are not similar to it/them.

Thus one gets, analogous to the colexifying term in (12a.), 'beards have got something to do with chins, but are not similar to them,' and so on. Note that the 'is like'-test yields negative results in contiguity-based semantic relations; in other words, the two tests most of the time yield mutually exclusive and unambiguous results (though see below for residual cases). The present approach avoids the fuzziness inherent into any conception of contiguity in that a semantic relation that is said to be characterized by contiguity is simply defined by yielding a positive result in the 'has got something to do'-test.

This yields a very basic and very broad distinction of semantic relations into two types, traditionally called similarity-based or metaphorical and contiguity-based or metonymical, which are both widely employed notions with a long historical tradition. The advantage of the present approach is that it combines these traditional notions with well-defined tests of the type used in lexical semantics to base the analyses on. If one or more of the constituents are verbal in nature, the test frame has to be accommodated by forming a gerund and inserting it into the test frame (as is done in Cruse 1986: 139 for another lexical semantic test). Thus, Gurindji *tiwu-waji* 'fly-AGENT' = 'airplane' would be (positively) tested for contiguity by saying 'an airplane has got something to do with flying.' This example also illustrates how derivatives which contain only one lexical root are tested, namely by creating the following test-frame:

(18.) a. <lexical root> is like <meaning of term>
b. <lexical root> has got something to do/is like <meaning of term>, but is not similar to it/them

Summarizing, the two very basic test frames can be applied equally well to diagnose semantic relations between different senses of colexifying lexemes, the relation of the derivation base to the meaning of the derived term as a whole in derivatives, and the semantic relations encountered within compounds and other types of morphologically complex expressions containing more than one lexical root. Results are normally unambiguous (see § 3.6.2.4. for exceptions). This is a major difference to the Koch approach, where the reasoning how in each individual case, metaphor, metonymy, and taxonomy is identified is not made clear. As Koch and Marzo (2007: 283) explicitly state themselves, "[i]n the present research, it is the linguist who, in the end, makes the decisions on whether there is or is not a motivational relation between two lexical units (and which is the formal and which is the cognitive relation at issue)" and go on to state that "[i]t would be tempting and empirically more sound to get this information from speakers." Ultimately, of course, in the present framework it is still the analyst who makes the decision whether the test yields a positive or negative outcome, but importantly, guided by the test frames. It still seems that test frames as employed presently are an important, albeit only first step, on the way to achieving intersubjectively comprehensible accounts. The general problem of intersubjectivity is remarkably not discussed in detail in the literature (Kilgarriff 1997): different analysts or informants may have diverging opinions as to the acceptability of the statements resulting from the tests.[28] Ideally, the test frames would be applied by using questionnaires translating the frames into the target languages and have native speakers evaluate the acceptability of the resulting statements (see, notably, Marzo and Rube 2006 and Marzo et al. 2006 for an application of the framework by Koch and colleagues to speaker-judgment based experiments).

[28] However, note that this problem also pertains to analyses not guided by test frames, and is perhaps even more acute here. Contrast two analyses of the association between 'the pupil of the eye' and 'little girl,' 'child' and like meanings (Tagliavini 1949, Brown and Witkowski 1981, Urban forthcoming, and Appendix E, 130): this pattern, according to Blank (2003: 55), "is explainable as the small reflection of oneself in the other's eye. We have, thus, an inseparable combination of similarity (ONESELF – THE LITTLE PICTURE) and conceptual contiguity (THE LITTLE PICTURE – THE ORGAN ITSELF)." In contrast, Riemer (2010: 376) says that "[t]his can be explained by metonymy. Our eyes have 'pupils' because of the small doll-like image that can be observed there: spatial contiguity, in other words, underlies the shift."

3.6.2.3. Refining analyses

Having established two basic types of relations, which are also implicitly or explicitly present in traditional analyses of noun-noun compounds and Cognitive Linguistics, they may be refined by asking: *in what way* are the two meanings inserted into the test frame, if a similarity-based relation is diagnosed, like each other, and, if a contiguity-based relation is diagnosed, *what* do the two meanings have to do with each other? These questions lead to the establishment of subtypes of both similarity-based and contiguity-based relations. At this point potentially arbitrariness comes into play, as this refinement can be done at different levels of granularity, and waterproof test frames to justify more fine-grained distinctions still need to be worked out. Concomitantly, contrary to the basic split into contiguity vs. similarity-based semantic relations, it is not entirely clear that these analyses in fact mirror linguistic reality in any meaningful way, although they offer a way of subcategorizing the observed patterns in an exhaustive and relatively elegant fashion. Therefore, these refinements are offered as parts of a preliminary typological grid that may serve as a starting point for further work on the topic, but, in the evaluative part of the present work, they will not be operationalized and semantic relations beyond the basic split as it has been so far established will not be tested for systematically.

For similarity-based relations, consider the following contrast between the following terms for 'beak':

(19). a. Laz *kinçi-çxindi* 'bird-nose'
b. San Mateo del Mar Huave *ombeay quiec* 'animal.mouth bird'

Both are similarity-based, because the 'is like'-test can be felicitously applied in both cases. However, the kind of perceived similarity differs: beaks are similar to noses in that they have a similar appearance (both are pointed etc.), whereas beaks do not look like human mouths, but are similar in function in that they are used for ingestion. Thus (19a.) can be called an example of a metaphor-based conceptualization of 'beak' that is driven by PERCEPTUAL SIMILARITY between the two referents, whereas (19b.) is a case of FUNCTIONAL SIMILARITY (see Geeraerts 2010: 34 for a distillation of early classifications of semantic change in which the same basic split appears to be recognizable). Kemler Nelson et al. (2003) provide empirical support for this distinction from psychology: they demonstrate that names children give to novel artifacts they are presented can be exhaustively described by the categories of perceptual and functional similarity.

The general semantic relation of contiguity can also be subclassified further. Given the fact that contiguity and the related process of metonymy are traditionally very broadly

construed (compare Bredin 1984 for a general overview and Waltereit 1999: 234 for an overview in terms of Cognitive Linguistics), a number of subtypes can be distinguished. Here, a division into as few subtypes as possible but as many subtypes as are required to classify the vast majority of cross-linguistically recurring semantic strategies as sampled for the present context is desirable.

First, contiguity, as traditionally and also presently conceived of, is often more specifically characterizable as SPATIAL CONTIGUITY, that is, the situation when the semantic relation between two meanings (i.e., what they have to do with each other, as defined by the contiguity test) is based on them typically or necessarily co-occurring in space. Representative examples are:

(20.) a. Badaga *gaḍḍa ~ geḍḍa* 'chin, beard'
b. Nez Perce *ʔipelíˑkt* 'cloud, thundercloud, sky'
c. White Hmong *qhov-ntswg* 'nose-hole' = 'nostril'
d. Aymara *uta* 'house, room'

Note that spatial contiguity includes part-whole relations, as seen in example (20d.) (as also stated by Waltereit 1999: 234 and Koch and Marzo 2007, among others). In contrast, the examples in (21.) are instances of FUNCTIONAL CONTIGUITY, i.e., the two meanings to be tested have something to do with each other in that one of them specifies the function or utility of the other:

(21.) a. Pawnee *rakaraaraaruukitaʔiituʔ*/rakaraa-raar-huukita-iit-uʔ/
'dishes-place-on.top.of-in.a.line-NOM' = 'table'
b. Chickasaw *aa-nosi-'* 'LOC-sleep-NMLZ' = 'bed'
c. Welsh *cysgod* 'shadow, shade, shelter'
d. Cheyenne *he'enénestôtse* /he'e-nén-hestôtse/
'female-nurse-thing' = 'nipple'

Terms in (22.) are instances of what will be called PERCEPTUAL CONTIGUITY. Characteristic for such terms is that their lexical designation highlights a particular aspect of the referent's appearance, such as color, shape, or its typical action.

(22.) a. Rendille *daáyto* /daáyi-to/ 'black-thing' = 'pupil'
b. Wintu *teˑd* 'blood, red'
c. Chukchi *ilə-lqen* 'damp-on.top' = 'swamp'
d. Ngambay *lò-ndùl* 'time-dark' = 'night'

As may have been noted, the perceptual vs. functional distinction is used in classifying both contiguity-based metonymic and similarity-based metaphorical lexical relations. This provides a desirable element of symmetry in the classification, but, crucially, it does not entail that the lines between contiguity and similarity become blurred. Contrast, for instance, (22a.) with Yanomámi *mamo ishiishi* 'eye coal,' which also means 'pupil.' The relation underlying the Rendille term is one of perceptual contiguity, that in Yanomámi clearly one of perceptual similarity. (22a.) highlights a particular aspect of the appearance of the pupil of the eye (its blackness) *directly*, but crucially, the pupil is not *like* blackness which would be a defining feature of a similarity-based conceptualization. In contrast, Yanomami *mamo ishiishi* explicitly compares the pupil of the eye with a piece of coal in that it is *like* coal *by virtue of* its blackness.

What will be called PROVENIENCE CONTIGUITY in the present study is intended as a cover term for what is usually called producer-product polysemy as well as some other types of lexical relations that have a common structure in that one of the tested meanings is the producer, source, or any other type of prerequisite for the existence of the second, such as lexical associations between the material an artifact is made of and the artifact itself. This category also covers the so-called "actual-potential polysemies" commonly found in the languages of Australia (O'Grady 1960, Dixon 2002: 56-57). Examples include:

(23.) a. Pawnee *haak- ~ rak-* 'tree, wood'
b. Wappo *húy* 'breast, milk'
c. Noni *kemfemtɛɛn* 'mucus, catarrh, flu'
d. Ngaanyatjarra *yawarra* 'wound, scar'
e. Nunggubuyu *aŋ̱bana* 'rain, raincloud'
f. Dongolese Nubian *šéma* 'wax, candle'
g. Cavineña *huaja* 'bee, honey'

CONFIGURATIONAL CONTIGUITY refers to a relationship between two meanings such that one is a certain configuration of the other. One possibility would be that one meaning denotes a mass of a certain substance and the other is a certain configuration of this mass, or an individual entity and a group of such entities. More generally, this category also includes sub-kinds of a given concept that are distinguished from the general meaning by virtue of being characterized by a distinctive property and thus by being a configuration of the more general meaning. In this sense, what is called configurational contiguity here draws near in some instances to the lexical relation of 'kind of.'

(24.) a. Pipil *a:t* 'water, river, rain, well, pool'
b. Mali *vutka* 'house, village'
c. Cashinahua *xau* 'bone, skeleton'
d. Greek *chártīs* 'paper, map'
e. Kapingamarangi *monowai doo* 'river fall' = 'waterfall'

The sixth and final subtype of a contiguity-based relation is TEMPORAL CONTIGUITY. This is mainly, but not exclusively, needed to account for semantic relations occurring in terms for phases of the day. Examples are in (25.).

(25.) a. Muna *alo* 'evening, night'
b. Abzakh Adyghe *śəble* 'thunder, lightning'
c. Copainalá Zoque *jama* 'sun, day'

3.6.2.4. *Residual cases*

While application of the 'is like'- and the 'has something to do'-test in the vast majority of cases delivers clear results, there is a residual class that cannot be unambiguously analyzed. One class of terms, which is widely discussed in word-formation textbooks, but occurs exceedingly infrequently in the terms for the meanings under investigation, are classical exocentric compounds, such as Hausa *sha ra'ba* 'drink dew' = 'calf (of leg).' Since the calf of the leg is neither similar to drinking nor to dew, nor does it have something to do with any of them in the strict sense, the tests do not take effect. However, since exocentric compounds like this in fact are based on a metaphorical comparison, but with the tertium not being realized by one of the constituents but lying outside of the construction, they are classified as being metaphorical in nature for present purposes.

For other cases, SEVERAL ANALYSES are POSSIBLE. This situation can occur due to two different reasons. The first is when colexifying terms have several different senses (i.e. more than two), at least one of which is analyzable in terms of similarity to at least one other sense, and at least one of which is due to contiguity to another one. Examples are in (28).

(26.) a. Ngambay *mùnjù* 'bean, kidney, heart'
b. Meyah *mei* 'water, river; sperm'
c. Chukchi *jiliil ~ jiləjil* 'tongue, language, blade of oar'

In (26a.), the obviously perceived similarity is the reason for using the same term for 'bean' and 'kidney' (compare English *kidney beans*). On the other hand, spatial contiguity may be involved in the conflation of the meanings 'kidney' and 'heart.' Likewise, with regards to

(26b.), a 'river' is a configuration of water (and therefore contiguous), whereas 'sperm' may be conceived of as being *like* water; and in (26c.), the 'tongue' stands in a functional relation to language in that it is used to produce sounds, whereas the 'blade of an oar' has got nothing to do with the tongue save for it being perceptually similar to it.

The second situation that leads to several possible analyses is when actually both tests yield positive results. Examples of colexifying lexical items for which this is true are:

(27.) a. Abipón *-aan- ~ -aanl- ~ -aana* 'thorn, needle'
b. Bezhta *baƛa* 'gut, sausage'
c. Kolyma Yukaghir *iŋd'i:* 'sinew, thread'

All examples can be conceived of as cases of colexification motivated by functional contiguity: a thorn may serve well as a needle by virtue of it being sharp, guts are used to produce sausages, and sinews may serve as a natural material to be used as thread. However, it is also at least imaginable that both senses are not in a functional relation, but are perceived to be similar to one another, and it would require a fair amount of ethnographic knowledge to determine whether, say, the Abipón used the thorns of plants as needles to substantiate this scenario.[29]

Finally, it must be mentioned that some apparent conceptualization strategies remain UNCLEAR. This pertains mostly to analyzable lexical items and reflects the fact that some denominations rely on highly culture-specific conventions that cannot be analyzed without intimate knowledge of cultural scripts and/or ethnographic information. A fine example is the Mali term for flood, *milatka avuouk*, which consists of lexical roots *milat* 'coconut shell' followed by the masculine singular suffix *-ka* and *uouk* 'grandmother,' preceded by the masculine singular possessive marker *av-*. The underlying metaphorical transfer pattern is at first glance probably unclear to most people who do not speak Mali, and indeed, this fact might lead someone who is not an expert on the Mali language to have serious doubts about the correctness of the morphological analysis of *milatka avuouk*. Stebbins (n.d.: 16), however, explains: "The scraper referred to here is a half coconut shell, used to scrape small weeds from food gardens. This expression makes use of the term grandmother to refer to something gigantic. For example, the cassowary is also known as the 'grandmother of the birds.' This type of flood is so powerful that it 'scrapes' away trees from the banks of the river" (see § 6.2.3.4. for a cross-linguistic survey of such kin-based

[29] Indeed, use of thorns as needles is documented, for instance for ancient Peruvian cultures (Harcourt 1962/2002: 9) and the Western Apache (Moerman 1998: 53). Still, this evidence is merely anecdotal in nature and does not rule out a similarity-based conceptualization in other languages.

metaphors). Obviously, it is an extremely lucky and rare coincidence that Stebbins happens to explain the underlying conceptualization of *milatka avuouk*. The conclusion that can be drawn from this example is that some cases where the semantic relation between constituent parts and overall meaning of the complex term in the database underlying the present work remains unclear, it could in principle be motivated if enough detailed cultural information were available that allow to "make sense" of the conceptualization.

On the other hand, the policy of allowing the relations between constituents and the entire complex term to be unclear also increases the danger of erroneous superimposition of morphological complexity when in fact none is there. For instance, given that in Yanomámi *u* is 'liquid' in the extreme case an analysis (not proposed here) of *puhutu* 'bud' (or any other word containing the vowel *u* for that matter) might be **p-u-h-u-t-u* '??-liquid-??-liquid-??-liquid.' A minimum of human judgment cannot be entirely eliminated here, in spite of the practical conventions for operationalizing the extraction of morphological complexity to be introduced in § 3.7.2.1.

Unclear conceptualization strategies are sometimes also encountered in colexification. As noted by Haiman (1974), it would be a quite remarkable coincidence if two languages had accidental homonyms with precisely the same meanings. These cases are therefore all the more interesting, and fleshing out the semantic motivation that links the two meanings is a challenge that would be worth pursuing. Two examples are in (28.).

(28.) a. Khalkha *xuvar* 'smallpox, flower, picture'
Kolyma Yukaghir *šörilə* 'flower, picture'
b. Sentani *ja* 'day, rain, already'
Bezhta *wodo* 'day, rain'

Note that in some of these cases a link is conceivable in that the semantics of the colexifying lexemes do not seem to be entirely unrelated, but its precise nature escapes analysis. With regard to (28b.) specifically, one would suspect, given the evidence of the present sample that some languages (Guaraní and Mandarin) colexify 'day' and 'sky,' while others (Katcha and Manange) 'sky' and 'rain,' that at an earlier stage, Sentani *ja* and Bezhta *wodo* also had the meaning 'sky' that served as a semantic bridge linking 'day' with 'rain.' However, Nikolayev and Starostin (1994) suggest that phonological developments in Bezhta led to the collapse of two originally distinct words with the meanings 'day' and 'rain' respectively.

3.6.2.5. *The notions of Semantic Domains and Domain-Supported Metaphor*

For the standard theory of metaphor and metonymy in Cognitive Linguistics, the notion of cognitive domain plays a key role, because for the differentiation of metaphor and metonymy, difference versus identity of semantic domain is the key differentiating feature (compare the definitions quoted in § 3.6.2.1.6.). A domain is understood in Cognitive Linguistics to be the sum of the background knowledge needed to understand the meaning of a particular linguistic item. In more formal parlance, concepts such as 'arc,' 'radius,' 'diameter' etc. are *profiled* against the background domain of 'circle' whose meaning is presupposed by the meanings of the profiled concepts (Langacker 1987a). Similarly, the days of the week can only be understood against the background knowledge of what a week is in the first place, and the meaning of week in turn recurs to a set of calendrical concepts such as 'month' etc. that are needed to understand the meaning of 'week' (Lakoff 1987).

Fine-grained subsequent research has argued, however, that a sharp division between metonymy and metaphor as strictly distinct mechanism along the lines of the standard definitions is hard to maintain and leads to analytic problems. In particular, it was argued that a metonymic base can be recognized for metaphorical processes (Goossens 1990, see also Barcelona 2000b for discussion). This gave rise to the so-called "demarcation problem," i.e. the question where to draw the line between metonymy and metaphor. Some scholars have called into question whether a strict division can be meaningfully maintained between mappings within the same domain and mappings that cross the boundaries of one domain, and some have therefore argued that it would be best to abandon the distinction between mappings within and across domains as defining properties of metaphor and metonymy (Feyaerts 2000, Riemer 2002b). Identity and non-identity of domain, in this view, would be maintained as a descriptive parameter, but would crucially be independent from the distinction between metaphor and metonymy.

Consider, as a first approximation, the following examples. These are instances of metaphors at the utterance level, which are discussed before turning to word-level metaphors because utterance-level conceptual transfers are the traditional subject of analysis in Cognitive Linguistics:

(29.) a. *As the sun sinks, the young bats stream from the cave-mouth like smoke and set off on the first stage of their long journey south.* (British National Corpus, F9F 641)
b. *Her words drifted like smoke.* (British National Corpus, GUK 2731)
c. *The fog seemed to part without warning, revealing the great headland of Rhuaival over three hundred metres above us, mist trailing like smoke from its peak and even a stray beam of sunshine lighting up the craggy outlines of the cliffs.* (British National Corpus, CRJ 899)

The source of the metaphorical transfer is in all cases 'smoke,' which belongs to the domain of aerosols in the sense used in physics, i.e. suspensions of fine solid particles or liquid droplets in a gas.[30] In the examples above, the comparison is already made explicit in that it is signaled overtly by a prepositional phrase headed by *like*. (29a.) and (29b.) are textbook examples of metaphor. But what about (29c.)? The comparison of 'mist' with 'smoke' is not an instance of any kind of contiguity: they do not necessarily co-occur together either spatially or temporally, they do not cause each other, they are not in a part-whole relation with each other, etc. Like (29a.) and (29b.), the relation between the two concepts would also be positively tested for metaphor by the 'is like'-test. If, however, it is accepted that 'mist' and 'smoke' belong to the same experiential domain, this is a case of metaphor that does not involve a conceptual trans-domain transfer. Moving on to the word-level, lexico-semantic associations between 'mist' and 'smoke' are not very common cross-linguistically, but they do occur, for instance in Jarawara, where *hote/hotone* is used with both meanings.

As an additional complication, there appears to be a gap between very technical definitions of what a domain is within Cognitive Linguistics itself. Croft (1993: 339) defines a domain as "a semantic structure that functions as the base for at least one concept profile," while Barcelona (2000b: 32) calls them "structured blocks of knowledge and experience which constitute the background for linguistic meaning," and the term indeed seems to be frequently used in a much looser sense that is more akin to the traditional notion of semantic field as opposed to cognitively oriented reasoning. Indeed, building the notion of domain into the definitions of metaphor and metonymy "raises the additional problem of stating precisely what an experiential domain is, when two domains are different, and when a domain is superordinate to another domain." (Barcelona 2000: 32).

Problems also arise when the technical and loose definition of domains clash. Heine (1997: 137), for instance, explicitly speaks of the "domain" of body-parts, and in discussing semantic shifts and extensions of body-part terms, aligns himself with the standard theory of metaphor as conceptual trans-domain mappings (1997: 139). Surprisingly, he then summarizes "that metaphor... is the only tool that takes care of the main features that characterize the transfer from object to body-part (e.g., from 'mouse' to 'muscle'), *from one body-part to another (e.g., from 'finger' to 'toe'),* or from body-part to inanimate part (e.g., from 'eye' to the 'eye of a potato')" (Heine 1997: 143, emphasis added).

For the present work, the notion of semantic domain is therefore employed, but in a relatively loose fashion, given that the goal is not to provide a cognitively or psychologi-

[30] Obviously, this scientific definition will not be present to some or probably even most speakers, but that concepts such as 'mist,' 'smoke,' 'steam,' 'ash clouds,' etc. do form a coherent semantic domain is suggested by the strong lexico-semantic ties between these meanings cross-linguistically (see § 6.3.2.2.).

cally sound account of semantic classification. Reference will be made to the domain of artifacts, the domain of body-parts and so on, bearing in mind Brinton's (2000: 112) definition of a semantic field as "a segment of reality symbolized by a set of related words [that] share a common semantic property." Consequently, the standard Cognitive Linguistics account of metaphor and metonymy is departed from in that the notion of semantic or experiential domains are not taken to be defining properties for these conceptual operations, but rather, the present framework relies exclusively on the established test frames to diagnose the presence of either. As the fairly detailed discussion of 'smoke' and 'mist' has already made clear, intra-domain metaphorical mappings are explicitly allowed (compare Rice's 2012: 35 metaphor PARTS ARE OTHER PARTS explained as "intra-domain metaphorical mapping based on form similarity" to describe denominations such as *setth'utthila* 'my nipple,' literally 'my-breast-head-hand(extremity)' in Dene Sųłiné). Consider also the body-part and body-fluid terms in (30.) and (31.).

(30.) a. Bakueri *ikéngé já ḿmɛ́ndɛ́* 'neck GEN leg' = 'ankle'
b. Efik *i'nua* 'mouth, nipple' (inter alia)
c. Swahili *shavu la mguu* 'cheek of leg' = 'calf'

(31.) a. Nez Perce *simqéheqs* /símqe-heqes/ 'penis-pus' = 'semen'
b. Abzakh Adyghe *pe-šən* 'nose-pus' = 'viscous snot'
c. Kwoma *moku sobo* 'semen raw/unripe/pure' = 'urine'

However, the notions of source and target domain are borrowed from Cognitive Linguistics because of their descriptive usefulness by referring to the element of a similarity-based complex lexical item and that of a contiguity-based complex lexical item triggering the tests to become positive as SOURCE CONCEPTS and to the overall meaning of complex terms as TARGET CONCEPTS. The term is used also for patterns of colexification where a particular direction of mapping suggests itself, but has to be taken with a grain of salt here, because, in spite of claims in Cognitive Linguistic literature to the effect that the source domain is in the overwhelming majority of cases concrete and the target concept more abstract (e.g. Kövecses 2002), directionality of semantic extension is not a priori clear in the absence of overt marking by derivational processes (Koch and Marzo 2007, Umbreit 2010, Urban 2011, see also Grady 1999 for a critical view from within Cognitive Linguistics).

It is probably not surprising that Wilkins (1996), a study exclusively devoted to semantic change in the domain of body-part terms, also allows for domain-internal metaphorical processes. Examples such as those in (30.) and (31.), which are all similarity-based and therefore metaphorical in nature, lead to the recognition of a subtype of

similarity as defined above, namely similarity that is at the same time assisted by the similarity being established between two referents in the same domain, that is, in other words, DOMAIN-SUPPORTED. Metaphorical transfers are powerful, but bold conceptual operations, and it appears possible that intra-domain transfers provide an additional element of conceptual fastening within the original domain to maintain comprehensibility of the semantic extension.

3.6.2.6. Contiguity Anchoring

Such conceptual bonding is also established by another subtype of metaphorical transfer which will be called CONTIGUITY ANCHORED and pertains exclusively to morphologically complex lexical items. Consider again the data already presented in (1.), reproduced here for convenience as (32.).

(32.) a. Bezhta *beš* 'skin, bark'
b. Mbum *ŋgàŋ-kpù* 'skin-trunk/tree' ='bark'

(32a.), from Bezhta, is colexifying and does not feature any overt marking that would mark the metaphorical connection between the two meanings of the term. (32b.), in contrast, is identical to (32a.) with respect to the semantic transfer that occurred (metaphor, based on resemblance of the concepts 'skin' and 'bark'), but formally, the term is a compound, with *kpù* 'trunk, tree' featured as an additional element. Now, 'trunk, tree' is contiguous to the meaning of the term as a whole, 'bark' (more specifically, there is a part-whole relation between the two meanings). This does not affect the overall metaphorical nature of the term as diagnosed by the 'is-like'-test. Therefore, it will be said that terms such as (32b.) are characterized by metaphor, that is, the semantic relations between the source and target concepts can be described as SIMILARITY that is CONTIGUITY-ANCHORED. In fact, most cases of perceptual similarity found in the data are of this type.

3.6.2.7. Why no taxonomic relations?

Koch (2001) and Koch and Marzo (2007) employ "taxonomic relations" as a matter of course without explicitly defining them. Indeed, hyponymy and hyperonymy are among the most well-established and best-known types of lexical relations. A distinction between hyponymy (defined in lexical semantics by test frames such as Xs are Ys) and the more narrowly defined taxonymy (defined by the more specific test frame Xs are a type of Ys) is often made (Cruse 1986, Croft and Cruse 2004), and it is unclear which of the two are meant in the Koch approach. Further, taxonomic relations as descriptive semantic relations are afflicted with a number of problems that are not immediately obvious. Cruse (1986: 137),

for instance, discusses some problems of taxonymy and presents seemingly straightforward examples which however do not produce normal results with the standard test frame for taxonymy (see also Cruse 2002 for a more complete outline of problems with hyponymy and a prototype-based approach to account for some apparently aberrant subtypes and Murphy 2003: ch.6 for review). Koch and Marzo's taxonomic super- and subordination appear to be very similar to, if not identical, with Cruse's taxonymy, but this is hard to be sure of as they do not offer an explication of these notions. One instance of taxonomic subordination Koch and Marzo (2007: 270) mention that does not seem so clear to the present author (as a native speaker) is German *stadt-rand* 'outskirts,' literally 'city-edge' which is said to be a taxonomic subordinate of *rand* 'egde.' Is there really a cognitively entrenched domain of 'edges' which has subordinates such as *stadtrand*?

The boundaries to both metaphor and metonymy (i.e. endo- and exocentric compound on the level of the signifier as far as compounds are concerned) appear to be more fluent than one would expect, although clearly taxonomic relationships are psychologically real at some level (see e.g. Moss et al. 1995).[31] Whether two signifieds are in a taxonomic relation with each other seems to vary to a considerable degree with the intuitions of different speakers/linguists (see Bright and Bright 1965 for discussion of difficulties with taxonomic relations in a fieldwork context).

The present approach therefore aims to minimize taxonomic relations as analytic categories altogether. In similarity-based conceptualizations, taxonomic relations do not arise in the first place because of the typically exocentric nature of such compounds, where the head constituent is not susceptible of being in a taxonomic relation with the overall meaning of the term. In contiguity-based conceptualizations, the test frame is designed to automatically target the element being contiguous to the target concept. In instances of colexification that might be thought of as exhibiting a taxonomic relation between their senses (which are rare anyway), the problem can often be avoided because, as Köveces and Radden (1998: 53) and Radden and Kövecses (1999: 34) have suggested, relations between some lexical items that could potentially be treated as taxonomic in nature can equally well

[31] Note that classic instances of exocentric ("bahuvrihi") compounds such as *egghead*, *redbreast*, etc., where the literal meaning of the compound denotes a salient aspect of the denotated entity of the term, interestingly are hardly represented in the slice of the vocabulary presently under investigation. These are typically said to be metonymic in nature (part for whole, see e.g. Ungerer 2002: 551). In some of these, metonymy and metaphor interact in often intricate ways which would pose an analytic problem, and, in fact, as Geeraerts (2002a) points out, several analyses for such items are possible. It would be interesting to further investigate just how common such compounds are, and in what areas of the lexicon they typically appear.

be accommodated under a broad notion of contiguity-based metonymy.[32] In the present framework, this would mostly be a relation of contiguity of the configurational subtype.

However, there are a number of recurrent patterns of colexification where indeed a taxonymic relationship between the two senses appears to be the only acceptable analysis. Notably, these are restricted to semantic areas that clearly form a coherent semantic field, and there is well-established evidence from prior research that indeed this semantic field is structured according to taxonymic principles. These are, first, the domain of ethnobiological classification, which is hierarchical in nature (e.g. Berlin 1992) and where colexification across the levels of the taxonomy are frequent (e.g. Berlin 1972) and second, the domain of artifacts (compare Cruse 1986: 147).[33] Examples include:

(33.) a. Imbabura Quechua *yura* 'tree, plant'
b. Huambisa *kuntin* 'bird, animal'
c. Miskito *raks* 'weapon, rifle'
d. Wappo *má·kina?* 'car, machine'

These types of colexification will be said to exhibit TAXONOMIC AMBIGUITY. This is closely related, if not in some cases identical, to the phenomenon of autohyponymy (Horn 1984, see also Becker 2002 for discussion relevant for word-formation). Note that there is the problem, at least for theories which seek to keep semantics and pragmatics distinct, of adequately distinguishing between context-triggered implicatures (which belong to the realm of pragmatics) and truly lexically entrenched senses (which belong to the realm of the lexicon) when meaning conflation of this type is concerned.[34] There is, unfortunately, no good principled methodological decision available to approach this issue. Given the fact that taxonomic ambiguity of this type plays an important role in the expansion of ethnobiological classification systems, at some point the semantic association must be viewed as belonging to the lexicon rather than being merely pragmatically conditioned. Moreover, given that the distinct senses made their way into dictionaries, it seems unlikely that these cases are pragmatically induced readings only, although of course the possibility cannot be altogether ruled out.

[32] Note also the group of conceptual transfers known as "generic is specific," which are argued to be metaphorical by Sullivan and Sweetser (2010) but considered metonymic by Rice (2012). Examples like these illustrate that there is little agreement in the cognitively oriented literature as to such cases.

[33] Note also that Rosch and Mervis (1975), while arguing for a prototype-based account of membership in the category 'furniture,' at no point deny that 'furniture' is a category.

[34] This distinction itself has been subject to criticism, see for example Nunberg (1979).

3.6.3. ISSUES IN COMPARATIVE SEMANTICS

3.6.3.1. Introduction

Cross-linguistic studies, in particular as far as semantics is concerned, require some important adjustments so as to render analyses meaningful and justifiable. Some of them were already noted; the next section is concerned exclusively with issues that arise when attempting to compare a large number of languages with respect to semantic features or, as is done here, semantic processes operating on the level of the lexeme. It is not without good reasons that recent approaches to cross-linguistic semantic analyses or *semantic typology* (e.g. Levinson and Meira 2003, Ameka and Levinson 2007, Bohnemeyer et al. 2007) operate with fine-grained data gathered using questionnaires with largely non-linguistic stimuli during actual fieldwork on relatively few languages. Indeed the question whether it is possible to arrive at meaningful results about cross-linguistic semantic patterns using published sources is in principle open.

Up to now, the discussion has operated with two assumptions the foundations of which require extensive additional discussion. Semantic relations between senses were posited by invoking the traditional notions of metaphor and metonymy, with the difference that in the present work these are established by test frames as used in lexical semantics. However, these test frames operate on the level of the metalanguage, not the object-language, which is on the one hand probably unavoidable. On the other hand, the fact that the locus of analysis is metalanguage glosses requires attention. First, the raw data on which this study is based is information from dictionaries, and, in the particular case of colexification, dictionary glosses for lexical items. Second, in the above classification, dictionary glosses were used as the basis for distinguishing different types of semantic relations, both in morphologically complex expressions and in cases of colexification. But it is not at all given that dictionary glosses correspond to senses as defined in lexical semantics! Further, taking into consideration simplex terms with several different meanings also introduces another problem. With complex terms, the analysis is in principle straightforward: one would, oversimplifying, have to establish the meanings of the constituent parts and state their semantic relations to the overall meaning of the complex terms. By contrast, without any overt mechanism, the analysis is much more complicated because there are no morphological or other clues to the different meanings of the terms in question. Among the associated problems with this general concern are (i) the notoriously difficult distinction between polysemy and semantic vagueness or generality, in particular (ii) in its relation to metalanguage issues and cross-linguistic comparison.

3.6.3.2. Polysemy vs. Vagueness

Polysemy is a notorious concept. While well-entrenched in the vocabulary of semanticists, it is probably the hardest of all lexical semantic relations to define unambiguously. Much recent research has dealt with the problem of delimiting it unequivocally from semantic vagueness. It has been demonstrated in great detail that traditionally employed tests to distinguish the two are problematic and often yield mutually contradictory results (Geeraerts 1993, Tuggy 1993, see Dunbar 2001 for a reply). To make matters even worse for cross-linguistic studies, these tests would have to be employed on a language-specific level for every individual case in every individual language. It is methodologically a no-go to make claims as to the status of a particular lexical item in a particular language with respect to polysemy and vagueness without careful application of the available tests (and even then, statements should be made with great caution). For instance, Terrill (2006) establishes that Lavukaleve *tau* which may translate to English as either *hand* and *arm* and might hence be dubbed polysemous is in fact semantically vague and could therefore be glossed with 'limb' rather than 'hand, arm.' The criticism of premature analyses in terms of polysemy expressed by Enfield et al. (2006: 141) and their conclusion that "[t]he burden of claiming polysemy is to explicitly establish it using linguistic tests" is well-founded. Likewise is Enfield et al.'s statement that "[s]tandard sources, such as dictionaries, do not provide the information required for distinguishing between a term's status as general or ambiguous." In a similar vein, Evans (2010: 524) points out that "not all sources have gone through the necessary analytic steps to demonstrate unquestionably whether monosemy or polysemy is involved." However, establishing the precise status of a given lexical item with respect to the distinction between ambiguity and vagueness is obviously impossible by sheer time restrictions, which require large-scale analyses such as the present one to be based on extant sources, such as dictionaries. Dictionaries, however, are in the first place practical tools, and their goal is not necessarily to provide detailed semantic analyses of its headwords, but rather to group distinct senses of a given lexeme in a meaningful way (employing the traditional lexicographic procedures concerning the microstructure of lexical entries, such as nesting) so as to facilitate the user's very purpose of consulting a dictionary: to find out about the meaning of a given word.

3.6.3.3. Metalanguage biases and the issue of sense division

Another issue that arises is the role of the metalanguage chosen for semantic analysis. As Malinowski (1935: 11) famously points out, "translation must always be the re-creation of the original into something profoundly different," and this statement is true no matter whether what is to be translated are whole passages of text or individual lexical items. The linguistic sign is irreducible in that its meaning cannot be described or explained other

than by recurring to linguistic signs, either from the same semiotic system or from a different meta-system chosen for analysis. This is the semiotic principle (Peirce 1932: 230-231). It poses serious problems for semantic analyses, because the semiotic meta-system, the metalanguage, is itself not neutral, but will potentially bias the analysis made of the object language in systematic ways (Goddard 1994, Goddard and Wierzbicka 2010, Evans and Sasse 2007: 68). To use a very simple example, if English is used as the metalanguage to describe the meaning of the Burarra term *murna*, it will be noted that the translational equivalent of *murna* in English is sometimes 'finger' and sometimes 'hand.' Since two different metalanguage terms are needed to adequately capture the meaning of the object language sign to be analyzed, the conclusion that one might reach is that *murna* is not monosemous, since it covers two related but distinct senses that are distinguished lexically in the metalanguage, or that *murna* is at least vague with respect to the two senses distinguished in the metalanguage. However, if the metalanguage is to be Ngaanyatjarra, another Australian language, then one metalanguage sign, *mara*, would suffice to adequately describe the meaning of Burarra *murna,* and the conclusion that might be reached under these circumstances is that *murna* is a simple monosemous lexical item. "[T]he delimitation of the number of word senses is always at the mercy of the metalanguage chosen for the analysis, and therefore open to potentially unlimited different analyses" (Riemer 2005: 124).

Recent work in lexical typology has lead to the realization of the momentousness and consequences of the metalanguage problem (e.g. Koptjevskaja-Tamm 2008: 43), has found a workable, albeit not optimal solution to circumvent the serious problems posed by the semiotic principle: instead of trying to characterize the semantic properties of a given set of terms that are to be compared on a language-internal basis (this would first involve an analysis of Burarra *murna* only using lexical semantic methods and establish its status with respect to homonymy, polysemy and vagueness) and its place in the language system emically, one goes for an etic characterization that "sets out all logically distinguishable possibilities regardless of whether or not individual languages group them together" (Evans 2010: 509). This approach is adopted also in François (2008), and since it is not based on language-internal tests, but on cross-linguistic comparison, the term COLEXIFICATION is used rather than POLYSEMY to reflect the different methodology. Since the present study operates on a similar basis, the terminological difference is adopted here. Colexification, as a comparative concept in the sense of Haspelmath (2007, 2010), is thus used to refer to any kind of conflation of several distinct meanings, defined as metalanguage glosses needed to capture the full semantic range of an object-language linguistic item. This does not entail any commitment as to the internal semantic structure of the analyzed terms and is to be understood strictly as a convenient cover term employed from a cross-linguistic point of

view. That is, with respect to the meanings 'finger' and 'hand,' one would simply restrict oneself to noting that there is at least one language, English, which employs two distinct lexical items, the meanings of which are conflated in Burarra. Using this approach, potentially valuable information for individual languages is lost. In particular, there is the problem of spurious sense division (Riemer 2002a): for instance, saying that Burarra *murna* conflates two meanings, 'finger' and 'hand,' conceals the possible elemental monolithic semantic structure this word might in reality have. The precise status of a particular linguistic item with respect to the monosemy-polysemy distinction would need to be established on a language-internal basis using lexical semantic tests (although these are plagued with problems themselves, as shown by Geeraerts 1993).[35] [36] Furthermore, Riemer (2002a, 2005) provides an account of semantic extensions that allows to circumvene many of the problems associated with the problematic distinction between polysemy and vagueness.

However, it is of utmost importance to note that the language in which testing for semantic relations is carried out is the metalanguage, not the object language itself for which the very relations are to be analyzed. Furthermore, it is not clear that the entities which are tested for the presence of this or that semantic relation (that is, dictionary glosses) indeed really are distinct senses in the object language. How can these unavoidable requirements (at least in the context of the present work) be justified and how can influences from the metalanguage on the analyses be excluded? The view adopted here is heavily influenced by the framework exposed in Riemer (2002a, 2005), and therefore Riemer's account will be described in detail in the following.

First of all, according to Riemer, it is not the case that the problem of potential influences of the metalanguage chosen for semantic analysis is genuinely associated with cross-linguistic work, but instead is a fundamental fact that all semantic work inevitably faces: "In fact, it is only in some translation metalanguage that the meanings can be brought to light and discussed in the first place: the only way we can talk about the different senses of a … lexeme is by providing a paraphrase of them in some different

[35] Compare also Sandra's (1998: 371) critique of certain types of analyses in Cognitive Linguistics, noting "the suspicious lack of a set of decision principles, which would make it possible to decide in an objective and replicable way whether two usages of a linguistic unit (lexical item or grammatical construction) are distinct or not" and summarizing that "[i]n other words, what is lacking from the enterprise is a set of scientifically valid principles."

[36] This approach is not anglocentric, just because English happens to make more lexical distinctions than Burarra. Similarly, English is lexically underspecified with respect to the referent 'lip' when compared with Dadibi, which distinguishes between *gani* 'lower lip' from *pedauwali* 'upper lip,' and with respect to 'lip,' English would be evaluated according to the same procedures that are applied to all other languages that are being investigated.

semiotic system" (Riemer 2002a: 5). This echoes the cardinal importance of the semiotic principle as formulated by Pierce (1932) for any attempt of semantic description. Riemer (2002a: 5) further argues "that metaphor and metonymy retain significant explanatory usefulness in spite of the restriction of their applicability to metalanguage glosses." Departing from the probably uncontroversial view that a linguistic item is a device for co-categorization of a variety of (sometimes highly dissimilar) real-world referents that establishes an equivalence between them by the very fact that they are referred to by the same expression, Riemer (2002a: 5-7, 2005: 159-161) argues for a threefold distinction of levels on which this co-categorization may take place:

> Micro-level categorization is the process involved in the ordinary, unmarked use of a linguistic expression for typical tokens of its class of referents, for example, the use of the word *flower* to refer to a particular individual flower manifest to the speaker for the first time. This level of categorization is the site of what could be called 'micro-polysemy' of words, that is, the potentially infinite minute differentiation to which referents and the nuances which accompany them are open while still counting as typical members of the lexical category in question ... The micro-level of categorization is essential to speakers' ongoing ordinary use of language to refer to the world. ... At the opposite extreme, macro-level categorization is the process in which atypical referents are assimilated to a pre-existing lexical category. This is the domain of many linguistic phenomena, including irony, exaggeration and other types of rhetorical effect, and, in particular, many types of consciously employed metonymy and metaphor. ... In contrast to the unconscious nature of micro-level categorization, here the use of a lexical item for an atypical referent involves a high degree of self-conscious, metalinguistic awareness, since it represents a marked departure from the typical referential norms of the speech community. ... These two extremes jointly define the residual intermediate level of lexical categorization. This level comprehends an array of disparate categorizations which are neither absolutely typical of the lexical item in question, nor absolutely atypical. Examples of the types of phenomena on this level would include dead metaphors and idioms, slightly atypical referents, and some of what ... we may call 'contextually modulated' categorizations. (Riemer 2005: 159-161)

Micro-level categorization, in this view, does not entail the creation of new word senses (otherwise a new sense of the word *flower* would come into being every time someone refers to a particular flower). Macro-level categorization, however, clearly does. Theories of metaphor, then, unnecessarily restrict the application of metaphor and metonymy as technical terms to the upper levels of categorization, while they can be equally well seen as being operative on the micro-level of categorization. For instance, when a real-world entity is categorized as a 'flower,' it is its resemblance, its similarity to the concept 'flower' that is the operative principle that governs categorization. "This connection, which is one of resemblance, is a metaphorical connection par excellance" (Riemer 2005: 164). In this view, analyzing a particular semantic relation as metonymical or metaphorical does not

necessarily entail postulation of distinct senses. As Riemer (2005: 166) further points out, "[t]he validity of postulating metaphorical and metonymic links between particular metalinguistic glosses can thus be guaranteed if these glosses identify features of referents that are salient on *one* of the levels of categorization (i.e. either the unconscious referential level, or the conscious conceptual one)." But how can this be guaranteed? This is indeed a difficult question on which the validity of the whole analysis crucially depends. Riemer approaches this issue by the introduction of so-called "S-Glosses" as a subtype of all possible metalanguage glosses that could be employed to describe the meaning of an object-language term.

S-Glosses, in his terminology, are glosses with "properties of referents which are sensorily manifest to speakers/hearers. Since these properties are at least perceived by speakers, they are part of the categorization process, if only on the micro-level" (Riemer 2005: 169-170). Consider Riemer's account along these lines of the semantics of Warlpiri *pinti*, glossed as 'skin, bark, peel,' which is worth quoting at length not only because this will serve to illustrate Riemer's approach, but also because terms with a similar semantics are in consideration in the context of the present work and have in fact already been mentioned above:[37]

> All three glosses are S-glosses: differences between skin, bark and peel are all clearly perceptible to Warlpiri speakers: these things, quite simply, all look different. Adopting a micro-level interpretation of the semantics of *pinti* would involve seeing the three glosses as not reflecting separately entrenched senses. Rather, the cocategorization of the three denotations would be explained by the semantic commonality between them: ... skin, bark, and peel are all *similar*, this similarity being captured by the superordinate description just given. Adopting a macro-level interpretation, on the other hand, would mean recognizing a correspondence between metalinguistic glosses and separately stored senses in the mental lexicon. In this case, 'bark,' 'skin' and 'peel' would each refer to a separately entrenched polysemous meaning of *pinti*, one of which would have to be taken as the core sense, with the others related metaphorically to it. Given that Warlpiri speakers can certainly perceive a difference between the referents named by the three glosses, the metaphorical link is part of Warlpiri speakers' linguistic knowledge *at least* referentially, on the micro-level of categorization. As previously noted, whether it is also part of their cognitive representations of the semantics of *pinti* is a question that will ultimately only be made meaningful if clear brain correlates are identified for the notion of a separate sense. ... The glosses of object-language words, the statuses of 'core' and 'extended' attributed to them, and the metaphoric and metonymic links by

[37] Note in the context of this discussion also the example sentence for Kaingang *far* in the consulted source for this language: *nén kar vỹ, far nũ: ka kar, nén tánh kar, ẽg mẽg kar, ẽg ke gé*. 'Everything has *far*: trees, plants, animals, us, too' (original translation: 'Tudo tem 'pele': árvoles, plantas, animais, nós também').

> which the glosses are related, have to be interpreted quite strictly as theoretical terms within a metalanguage and not necessarily revealing the status and interrelations between different senses in a psychologically realistic way for Warlpiri speakers. The division of the glosses of each word into 'core' and 'extended' meanings that will be made in the analysis in this book is therefore not to be interpreted as claiming that the different metalanguage senses attributed to a word all correspond to different polysemous senses. Rather, the status of these senses as either separate meanings on the macro-level of categorization, or as 'modulations' of the same meaning ... on the micro-level, is left unspecified. The interpretation retains a minimal degree of psychological plausibility ... if ... only S-glosses are chosen ... Clearly, this will only be possible for words referring to 'concrete' or perceptually available entities (Riemer 2005: 171-173).

This, importantly, is exactly the case in this study, which is restricted from the very start mostly to perceivable "objects" in the extra-linguistic world for precisely the reasons for which Riemer finds it necessary to ground analyses on what he calls S-Glosses.[38] This reasoning can be illustrated with two examples, one metaphoric, the other metonymic. The first was already briefly mentioned: In Yélî Dnye, *mbu* is used to refer to mountains, but also "to conical elevations of any size, even a heap made by a burrowing crab on the beach" (Levinson 2008: 261). Given that conical shape seems to be the semantic feature most prominently encoded by *mbu*, the correct analysis on language-internal criteria would perhaps be that *mbu* is monosemous rather than polysemous with several distinct senses such as 'mountain,' 'heap made by crab,' etc. But note that the distinction between monosemy and polysemy is not at stake presently. What matters is that there is a relation of similarity between mountains and heaps made by crabs precisely because of their common conical shape. Whether this similarity-based grouping of referents under the umbrella of one lexical item occurs at the unconscious micro-level of categorization or at the conscious macro-level of deliberate linguistic behavior that is the domain of most Cognitive Linguistics analyses, or occurs at the intermediate level is a question that no commitment needs to be made to. The similarity relation between the disparate referents of *mbu* is enough to diagnose its senses as being in a metaphorical relation to each other.

A second example is provided by conflation of the meanings 'thunder' and 'lightning' into one lexical form, which is common in languages of South America (see Appendix

[38] Note that the view expressed by Riemer that 'skin,' 'bark' and 'peel' "quite simply, all look different" as well as the very notion of 'objects' in the extra-linguistic world entail some implicit assumptions which are at the very core ultimately philosophical in nature. Lyons (1977) is one of the few to explicitly state that the basis of his semantic views is a philosophical position of 'naïve realism' which he considers 'harmless.' However, it is important to note that taking a harmless point of view nevertheless still *is* taking a point of view.

E, 35,64), but occurs also in German *Gewitter*. The two meanings are in a relationship of contiguity with each other: they typically co-occur spatially and temporally. If one adopts a micro-level analysis of terms that can be translated as both 'thunder' and 'lightning,' one would assume that such terms are essentially monosemous, the general label reflecting their co-occurrence. On the other hand, a macro-level analysis would entail the assumption that there are two distinct lexically entrenched senses of the term, 'thunder' and 'lightning.' However, the explanation for this grouping of colexified senses would remain essentially the same. This account of semantic extension is not only descriptively elegant for cross-linguistic purposes. By remaining non-commital as to the level of semantic extension and hence to the issue of sense division, it also allows to circumvene to some extent the need to rely on an account of semantics in which lexical items host distinct reified senses which is both descriptively difficult to handle and, perhaps even more importantly, glosses over the dynamic nature of the emergence of semantic structure in the context of utterances (Geeraerts 1993, Allwood 2003).

3.7. PRACTICAL PROBLEMS

This section is concerned not so much with problems of a theoretical nature, but rather with questions that arise when actually extracting data from extant sources. A number of issues arise, some general in nature, others pertaining to either the formal or semantic side of the lexical items gathered only.

3.7.1. General Considerations

Data were excluded from quantitative analysis when the lexicographer explicitly marked a certain term, or its morphological make-up, as dubious. The same was done when lexicographers made explicit that a certain term is very rare or archaic, or that it belongs to a specific register of the language (this would include motherese, avoidance registers, etc.) and does not occur in ordinary speech.

Pawley (1996a: 189) points out that "there is a continuum, rather than a sharp division, between nonce forms and highly conventionalised expressions," and there is no principled fashion in which a lexical item's status with respect to this continuum can be inferred from dictionaries. Therefore, in the most extreme case, there is the potential issue that complex items listed in dictionaries might be artifactual, that is, neologisms coined in the course of the linguist's work on the language to refer to a specific entity in the non-linguistic world that does not constitute a part of the ordinary vocabulary of the language in question. Haspelmath and Tadmor (2009b: 10) discuss the case of Indonesian, which lacks a word for 'pasture' and speakers consequently make use of a complex term when urged to

name a word for this particular referent. Similarly, Sawyer (1965) remarks that the complex Wappo term *ʔùču·aʔ-meʔ hín* 'night-belonging.to sun' for 'moon' is "possibly contrived." While this seems unlikely in this particular case given the overall areal distribution of complex terms for 'moon' (see Appendix E, 38), the issue remains problematic in general. Further, overly long terms are suspicious in this respect. For instance, the consulted source for Yoruba gives *ibiti o dabi ẹnipe ilẹ̀ ati ọrun pade*, literally "place where land and sky appear to meet," as the Yoruba equivalent for 'horizon,' and indeed, Joseph Atoyebi (p.c.) informs that this is not a lexicalized expression in Yoruba but rather a circumlocutory definition of the concept. However, unless additional information such as that provided by Sawyer is given in the consulted source, there is no principled criterion available to sort out such cases when working with extant sources. On the other hand, it seems rather unlikely that there are very many of such non-lexicalized expressions in dictionaries, although a small bias in one direction or the other cannot be entirely excluded.

3.7.2. PRACTICAL PROBLEMS IN FORMAL CLASSIFICATION

3.7.2.1. Problems with extraction of morphological complexity

In principle, whether what one is dealing with is a morphologically complex lexical item can be determined straightforwardly. However, in practice, two problems arise.

The first relates to the way lexicographers choose to include information about morphological complexity. In the ideal case, constituent parts of complex items are directly given as part of the lexical entry. This is generally the case with data from Haspelmath and Tadmor (2009c), which, however, have the drawback that they are not so explicit about colexifying structures. The second best situation for the context of the present investigation is when dictionaries do not provide object-language constituent parts, but provide a "literal" translation which can then be used to infer the gloss of a given complex lexeme and to identify the lexical material present. Unfortunately, this is not always the case either. In the worst of cases, no information at all is included in lexicographic sources about possible morphological constituents of listed headwords. As a rule, where possible, grammatical descriptions of each sampled language were consulted, either in the form of separately published grammars or, where these were sufficient, grammatical sketches provided in the dictionaries themselves, to get an idea of the language-specific ways word-formation in each particular case and to get an overview of the form and function of derivational morphemes (if any are present). Further, orthographical conventions can be of some help here, e.g. when one or more constituent parts are written as separate words, which can then be looked up under own their individual entry. Often, especially as the investigation proceeded past its initial stages, regularities of

semantic association emerged which were in many cases of help for the identification of constituent parts (compare Brown's 1999 "educated guess approach"). In the database, it is coded for each individual dataset whether the morphological constituents of analyzable terms were explicitly stated in the source or whether they were inferred. Errors cannot be excluded, and are indeed likely to be present in one case or the other, either in the form of a complex formation not being recognized as such, or in the form of erroneous superimposition of morphological complexity that is in fact not there, or in correctly diagnosing morphological complexity, but erroneous recognition of constituents. In terms of quantity, this problem is much more likely to cause an underestimation of the degree of overtly marked items than an overestimation.

Second, a recurrent and unfortunate problem is that, for a given lexical item in a given language, some constituent parts are recoverable from the source, i.e. the lexical item is putatively a complex one, but not all are. This may be due to three different reasons: (i) the involved putative morpheme(s) are "cranberries" and as such due not have a lexical entry of their own in their source, (ii) the involved putative morpheme(s) are indeed meaning-bearing, but for some reason were not included in the consulted source and thus cannot be glossed, or (iii) the elements in question are altered by either synchronic morphophonemic rules of the language in question or by lexicalization (in which case the procedure outlined in § 3.7.2.2. applies). The policy adopted for such cases is the following: putatively complex terms were only counted as such if and only if (i) the number of identifiable morphological elements outnumber the unidentifiable ones and (ii) from the meanings of the identifiable morphological elements, a reasonable picture of the semantic relation between them and the meaning of the term itself can be obtained. For illustration, examples in (33.), all from Biloxi, were coded as being semianalyzable, and examples in (34.), again from Biloxi, were coded as being analyzable in spite of containing unknown elements.

(33.) a. hadi´xtciya^{n}´ /hadi´xi-tciya^{n}´/ 'urine-??' = 'bladder'
b. *nḳa-´toho* '??-lie.down' = 'bed'

(34.) a. *pĕ´xĕno^{n}n´ sŭpi´* 'fire:?? black' = 'coal'
b. *a´diṭo^{n}ni ~ adito^{n}´*/a´-ti-ṭ-o^{n}ni/ '??-eat-??-make' = 'table'

Exceptions from this rule were made only if the morphological constituents of complex terms are unclear, but the lexicographers explicitly provide a "literal" translation. For instance, the Lakhota term *wakȟiyątųwąpi* 'lightning' is not amenable to precise morphological analysis on the basis of the consulted source, but is said to be "literally" translatable as

"the thunderbirds are looking." Such statements (which are rare) were taken to be equal to identifiable morphological complexity for evaluation, though languages with sources where they are the only clue to morphological complexity were if possible avoided (which is why Lakhota is not in the core sample).

3.7.2.2. The problem of lexicalization

A related question arises if constituent parts can be inferred, but these are phonologically altered in comparison with their shape when occurring as free-standing forms. In the absence of any objective criteria to ensure that analyzable terms are also transparent in the senses defined above, such phonological deviations are taken to be a clue that the lexical item in question is LEXICALIZED. The term lexicalization has many related and unrelated uses in linguistics (Brinton and Traugott 2005). Here, by lexicalization is meant the diachronic process that causes the originally present morphological make-up of complex terms to lose psychological reality to speakers (they become non-transparent). This will subsequently, in some cases facilitated by high frequency of usage, blur morpheme boundaries up to the point that originally complex structures can only be restored by etymological work.[39] An example of this frequent diachronic process is Spanish *comercio* 'trade,' which goes back to Latin *commercium* with roughly the same meaning. In Latin, the term is synchronically analyzable, consisting of the prefix *con-* 'with, together,' *merx* 'merchandise, good' and an abstract derivational suffix. Obviously, there are no synchronic rules of Spanish phonology that would allow relating *comercio* to underlying *commercium*. Further, the fact that the Spanish cognate is clearly univerbated is evidenced by the loss of *merx* in the course of the development to Spanish. Now note that in Latin *commericum* itself there is a phonological mismatch in the final consonant of the prefix *con-*, which is assimilated in its place of articulation to the following consonant in the complex term. The crucial difference is, however, that this is due to a regular synchronic assimilation process of Classical Latin grammar, and is thus predictable.

For the determination of whether a given lexical item whose status with respect to analyzability is dubious, the guideline adopted is the following: whenever the phonological deviations are accountable by synchronic phonological rules occurring at morpheme boundaries (such as assimilation, as in Latin) of the analyzed language, such terms are judged to be analyzable (but not necessarily transparent), whereas any alternations that are not are taken to be indicative of lexicalization, and lexicalized terms are as a rule not

[39] This process is also called univerbation by some authors.

considered to be analyzable from a synchronic point of view. As examples, consider data from Kashaya in (35.)

(35.) a. *muśa·laqol* /muša·la-ʔahqol/ 'snake-tall' = 'rainbow'
b. *qʰaṭʰuʔul* /ahqʰa-ṭʰuʔul/ 'water-old' = 'swamp'
c. *hokare·ta* /ʔoho-kare·ta/ 'fire-wagon' = 'train'
d. *qʰa·moš* /qʰa·-moš/ '??-sour' = 'star'

Examples (35a - c.) illustrate the regular phonological process of aphesis, i.e. loss of morpheme-initial syllables containing /ʔ/ or /h/ in polysyllabic words when entering into a construction, such as a compound (Oswalt 1961: 305).[40] (35d.) looks similar, but in fact, the first constituent, *qʰa-*, is, according to Oswalt (n.d.), "apparently derived by irregular contraction from *qʰa·ʔa* nightlong or *qʰaʔa* morning, producing an unpredictable accent on a long syllable." Consequently, (35a. - 35c.) are analyzed as being analyzable, whereas (35d.), although the morphophonological resemblance to free-standing forms whose semantic would match is intriguing, is said to be semianalyzable.

A more systematic and extreme case is posed by the so-called "combining forms" of nominal roots in Sora. This process is of great importance to the present study because "Sora makes extensive use of root/stem compounds and lexicalized derivational elements in the creation of its nominal lexicon" (Anderson and Harrison 2008: 321). In Sora, there is a systematic alternation with respect to the shape of nominal roots, depending on whether they occur bound or as free-standing forms. Examples are in table 2.

Free Form	Combining Form	Meaning
aŋgaj	*gaj*	'moon'
daŋgo	*daŋ*	'stick'
kinad	*kad*	'crab'

table 2: examples of free and combining forms in Sora, from Starosta (1992: 85-86)

At first glance, the combining forms appear to be the result of irregular contraction of an originally free form, somewhat reminiscent of the genesis of lexical affixes in North American polysynthetic languages. However, experts agree on the basis of comparative

[40] Interestingly, this process does not shut down transparency for speakers. When asked about the "literal meaning" of *qʰami lahwal* 'coast,' a Kashaya consultant answered "all I know is it's got something to do with water" (*ahqʰa*). Although the statement needs to be restricted to the individual case of Kashaya only, this piece of evidence corroborates the distinction presently made.

evidence that the combining forms are diachronically original, and the free forms are derived from them by a variety of morphological processes. Consequently, since the combining forms are primary, they are not treated as being lexicalized and are viewed as synchronically identifiable lexical elements when occurring in morphologically complex terms.[41]

However, in a considerable number of cases it is not possible to find the relevant information in the existing literature on the sampled languages. In these cases, the rule of thumb adopted is that if the deviation from what is expected if forms were fully transparent is no more than one segment, the respective analyzable term was considered to be analyzable rather than semianalyzable. Thus, for instance, Berik *tokwa es* 'spark' is considered to be analyzable, consisting of *tokwa* 'fire' and *ese* 'flower.'

3.7.3. PRACTICAL ISSUES AND METHODOLOGICAL ADJUSTMENTS IN DATA INTERPRETATION

3.7.3.1. Analyzability and Colexification

The descriptive categories of analyzability on the one hand and colexification on the other are not mutually exclusive, i.e., analyzable lexemes may sometimes have several senses, such as in the examples in (36.).

(36.) a. Comanche *puhihwi* /puhi-ekahwi/ 'leaf-shiny' = 'money, gold'
b. Itzaj *säk-b'ej* 'white/grey-trail/road' = 'Milky Way, highway'

Behrens (2002: 327), following Schmidt (1982), points out that "it is not entirely clear whether or not the homonymy-polysemy distinction is applicable to morphologically complex lexical forms (e.g. compounds)." Morphological analyzability as opposed to colexification is regarded as primary for the purposes of the present study, and consequently analyzable terms with several apparently distinct senses are categorized as

[41] A similar and somewhat problematic situation might obtain in Hani. For instance, *saqguq*, with *saq* meaning among other things 'muscle, meat,' is 'tendon,' while *sivqguq*, with *sivq* meaning 'blood,' is 'vein.' *Guq* alone is glossed as 'to need, to require,' which does not bear a recognizable semantic relation to either 'tendon' or 'vein.' Still, that the terms share this element raises the suspicion that really *guq* bears the meaning 'tendon, vein,' which is a common enough pattern of colexification (compare Appendix E, 141 and 147) but does for some (prosodic?) reason not occur as a simplex (note further that it occurs in *keelguq* 'Achille's tendon,' where one would expect 'foot' to be the meaning of the second element in a complex term if indeed *guq* bears the meaning 'tendon' and 'vein' alone; indeed, *aqkeel* means 'foot,' but again, only occurs in connection with *aq*). There are several similar cases. Since an analysis along the lines of the above, however, is unlike the Sora case conjectural, a more conservative solution is preferred, and, for instance, *saqguq* and *siyqguq* are treated as semianalyzable.

analyzable, not colexifying, for quantitative evaluation. However, in chapter 6 and Appendix E, patterns of colexification in analyzable lexemes are also borne in mind.

3.7.3.2. Idiosyncratic language-specific meaning extension

As François (2008: 163-164) points out, "the more languages we explore, the more examples we find of unique metaphors and unexpected cases of semantic shift – probably one of the most thrilling mysteries and charms of language discovery. But what generally happens is that we focus our attention on the most exotic cases, and overlook the information that is of most interest for the hunter of semantic universals: namely, that a great deal of lexical polysemies are in fact widespread across the world's languages, and, as such, deserve to be highlighted and analyzed." This is precisely one of the goals of the present study. But what to do with the unique cases of colexification that are encountered? Examples include:

(37.) a. Bakueri *e̱w̱umá* 'ball, orange'
b. Sahu *utu'u* 'root, aerial root, buttress'
c. Cayapa *ñi* 'seed, fire, flame'
d. Hawaiian *ānunue* 'rainbow, scallop-like design on tapa and tapa-beater'

In principle, it appears to be harmless and indeed appropriate to also take into account such unique conceptualization strategies. However, there are at least two problems associated with them. The first is that, if the semantic association by means of a purportedly colexifying lexical item is encountered only in one language, mere accidental homonymy is at least a possibility. This seems to be rather unlikely for cases such as (37a.) and (37b.) because the semantic motivation for the meaning conflation is quite obvious. However, more serious doubts, which are reinforced by the shortness in terms of segments of the object language term, about the semantic connection between 'seeds' and a 'fire' in (37c.). Indeed, recent work on comparative semantics, in particular work carried out in connection with the semantic map approach to cross-linguistic comparison has adopted an explicitly cross-linguistic stance to deal with this issue. In the words of Haiman (1974: 341), "[i]f a word exhibits polysemy in one language, one may be inclined, or forced, to dismiss its various meanings as coincidental; if a corresponding word in another language exhibits the same, or closely parallel, polysemy, it becomes an extremely interesting coincidence; if it displays the same polysemy in four, five, or seven genetically unrelated languages, by statistical law it ceases to be a coincidence at all." Croft (2003: 106) elevates this observation to a full-fledged typological principle for distinguishing polysemy from homonymy. It is adopted in the present work. That is, if a putative semantic relation is showcased only by one language in the sample, it is not taken into consideration for

quantitative analysis. This is an answer to the important question raised by Koch and Marzo (2007: 282) as to "from which threshold on are we allowed to disregard senses appearing in our material, as it has been defined by our methodology?"

It should be noted, however, that both Haiman and Croft are interested primarily in the semantics of closed-class grammatical items and recurrent cross-linguistic semantic overlap. The semantic range of closed-class items tends to be more restricted in the possible range of different uses than the comparably unconstrained members of the open-class lexical categories of a language's lexicon. Therefore, the criterion will undesirably sort out some likely cases of genuine metaphor-driven colexification such as (37a.) and (37b.), which are in principle no less interesting than cross-linguistically recurrent patterns of colexification. This leads to the second problem with unique cases of semantic extension, which is that at least some cases of unique patterns of colexification in the database are based on highly culture-specific connections, and it is therefore little wonder that they remain without a counterpart in another language. (37d.) is an example. Intuitively, one could image encountering another language in which the same word is used for 'ball' and 'orange' as in Bakueri, but one would not expect to find the same, or even a similar, semantic extension of 'rainbow' as in example (37d.) from Hawaiian. This illustrates that the boundaries between idiosyncrasy and rarity cannot, of course, be adequately drawn on the basis of the sample underlying this study alone. Therefore, it is safer to use Croft's criterion and draw a partly arbitrary, but not entirely unmotivated line in the sand by the application of this principle. In other words, the level of granularity of recognizing distinct senses is jointly defined by the glosses offered by the dictionaries in the sample languages.

For quantitative evaluation, however, unique conceptualizations that are manifested by morphologically complex lexical items are taken into account, because they can, at least in principle, be straightforwardly analyzed and are not plagued by analytic issues comparable to those arising with colexification. Furthermore, the above mentioned procedures with respect to the analysis of colexification also serve another important purpose, namely to normalize the raw data with respect to the level of detail in which lexicographers elaborate on different senses of a given lexical item. This may potentially bias analysis: a full-fledged dictionary of a "big" language will normally also provide more detailed information about the semantics of the individual lexemes listed in it, whereas a dictionary (not to mention mere wordlists) of a small minority language that has received comparably little attention by linguists will tend to be more coarse-grained with respect to the semantic distinctions that are recognized in the dictionary. This observation should not be read as a statement concerning the quality of the lexicographic work undertaken; nevertheless it is clear that socio-politically more important languages simply tend to be allocated more resources to produce comprehensive dictionaries. Importantly, however,

the more senses of a lexeme are listed, the more specialized they will be semantically, and the less likely they therefore are to recur cross-linguistically. For instance, Lessing's (1995) dictionary of Mongolian provides an impressive level of detail in the description of the semantics of its headwords. Consider the examples in (38.)

(38.) a. *ceceg ~ cicig ~ seceg* 'flower, smallpox, comb of cock, club (in cards)'
b. *terge(n)* 'vehicle, cart, wagon, carriage, car, rook (in chess)'

The procedure of only recognizing a pattern of colexification if it occurs in at least two languages sorts out the meanings 'smallpox,' 'comb of cock' and 'club in cards' of Khalkha *ceceg* and the meaning 'rook in chess' of *terge(n)*, and thus allows to normalize the typological data in a consistent fashion. This is desirable, because otherwise elaborateness of the source would lead to an unjustified increase in the amount of colexification diagnosed in the presently investigated slice of the lexicon when compared with languages where the source is more restricted. However, such influences cannot entirely be ruled out (see § 3.7.3.4). After all, the more senses a source provides, the more likely it is for one of them, however specialized semantically it may be, to occur elsewhere in the world.

3.7.3.3. Assessing Influences of different types of sources statistically

Influence of the nature of the consulted sources was in addition assessed for two of the major variables surveyed in this study: degree of morphological complexity and degree of colexification (recall that the relation between form and meaning assumed here is many-to-many. Thus, for instance, if a language has two terms for a particular meaning, one morphologically complex and the other not, this language is assigned a value of 0.5 analyzable terms for that particular meaning). To this end, different breakdowns of the statistics sample were used and then checked for statistical correlations. The following breakdowns were tested:

(i) The main source is or is not a vocabulary contained in the World Loanword Database

(ii) The main source is or is not a (short) unpublished manuscript, an appendix in a grammar rather than a full-fledged dictionary, or any other type of publication which is not designed mainly to provide lexicographic information about a particular language. The latter group includes, for instance, the etymological dictionaries for Chukotko-Kamtchatkan and Yukaghir (Fortescue 2005, Nikolaeva 2006) from which the Chukchi and Kolyma Yukaghir data mainly stem.

(iii) The main source is or is not a dictionary produced by members of the Summer Institute of Linguistics
(iv) The publication date of the main source consulted

The obvious possibility of using the number of pages the source has as a statistical predictor variable was refrained from, because sheer number of pages depends heavily on typesetting conventions and does not necessarily represent a fair measure of the overall scope of the source.

With regard to categories (iii) and (iv), no difference for any of the variables was obtained that could be meaningfully interpreted. Thus, there is no appreciable effect of the age of the consulted source on any of the variables under study, nor is there an effect of the provenience of the source. Concerning factors (i) and (ii) however, there is a statistically significant effect of the source type on the behavior of the language: in spite of the countermeasures described in the above section, there is an effect on the measured degree of colexification in the statistics sample that is caused by the source being a vocabulary contained in the World Loanword Database ($W = 442$, $p = .0407$, Wilcoxon rank sum test). Similarly, the measured degree of colexification is significantly lower for sources other than standard dictionaries ($W = 511$, $p = .001759$, Wilcoxon rank sum test). At least the second test remains significant when correcting p-values for multiple testing using the Bonferroni correction at $p = .014072$.

In effect, while there is no indication that the measured percentage of morphologically complex terms fortunately depends on the nature of the consulted source, the diagnosed amount of colexification apparently does to some extent. In a way, this result is hardly surprising: the purpose of the World Loanword Database simply is not to provide a highly detailed picture of the lexical semantics of each individual lexical item in the vocabularies (though, of course, there is some information). Likewise, the main job of an etymological dictionary is to identify cognates in genetically related languages and perhaps provide a reconstruction of the proto-word, but not to analyze in detail the semantic microstructure of daughter-language lexical items. Because of these results, there will be no following discussion and quantitative search for correlations that might trigger the measured degree of colexification. But this does not at all entail that the gathered data on covert semantic patterns in the lexicon cannot be put to meaningful use. It can, and it will be, for instance in the search for areal and universal lexico-semantic associations carried out in § 6.4.3.

Furthermore, there is an effect on the measured number of semianalyzable terms when comparing data from the World Loanword Database, in which information on analyzability and semianalyzability is coded directly by experts, to data from other sources.

The difference is significant at $p = .009424$ (W = 477, Wilcoxon rank sum test). Consequently, other data are likely to systematically overestimate the amount of truly semianalyzable terms and concomitantly to underestimate the number of analyzable motivated lexical items, which could not be identified as such. However, since this effect pertains to the majority of the datasets in the statistics sample and thus more or less to the sample as a whole, drastic effects on the conclusions drawn from it are not to be expected. However, it should be emphasized that when a term is called "semianalyzable" anywhere in the following chapters, this should always be read as "semianalyzable on the basis of the consulted source," thus not excluding the possibility that the term, given more knowledge on lexicon and grammar of a particular language, would turn out to be fully analyzable.

3.8. CHAPTER SUMMARY

This chapter developed a framework for the study of lexical motivation. It departs from previous approaches, in particular Koch (2001) and Koch and Marzo (2007), in both its formal aspect as well as in its semantic side. As for the latter, it is, unlike previous approaches, based on lexical semantic test frames to guide the analyst in his decisions as to classification. These yield a basic two-way split of semantic relations along with the establishment of further optional subtypes. The basic split is compatible with the notions of metaphor and metonymy as traditionally conceived of. In the course of the exposition, two methodological problems in the classification of semantic relation and the (preliminary) solutions chosen for them here were pointed out: first, the non-neutrality of any (natural) meta-language chosen for analysis and the concomitant issue of appropriately identifying senses in colexification (the problem of sense division), and second, the appropriateness of postulating semantic relations in the case of colexification.

The major variables that emerge from the discussion so far are the (i) relative degree of morphologically complex expressions, (ii) their type (lexical, derived, or alternating), and (iii) the amount of semantic associations by colexification. In addition, there is (iv) the variable of the relative proportion of metaphor- as opposed to metonymy-driven semantic patterns. However, since there are biases in the data with respect to the relative degree of colexification itself which are caused by the nature of the consulted sources, it will not be evaluated in quantitative terms. The relative degree of metaphor as opposed to metonymy, however, is evaluated using data from morphologically complex expressions as well as those of colexification, since, as shown above, this distinction is equally applicable to both types of lexical motivation.

Chapter 4

The Formation of Complex Lexemes: Types and Strategies

4.1. INTRODUCTION

The aim of this section is to give an overview of the impressive variety of different types of morphologically complex nominals that can be found in the world's languages, and thus to give a "feel" for the data to be analyzed in a more quantitative fashion in the subsequent chapter. At the same time, a subsidiary goal of this section is to introduce and explain the basic statistical techniques of data analysis that will be used in doing so. Since this study is both interested in areal as well as universal typological trends and correlations, what is required is (i) a method to assess areal patterns in a meaningful way and (ii) statistical tools for determining valid typological trends, that is, trends which are precisely not due to areal tendencies in the data. These techniques will play a major role in the next chapter which is concerned with a quantitative analysis of the gathered data. This section, on the other hand, introduces these techniques embedded in a more traditional typological discussion in order to not overwhelm the reader with a lengthy technical statistical discussion at one point in the thesis, but rather to introduce relevant methods of analysis in a piecemeal fashion to gradually familiarize him/her with them.

With respect to the main goal of this section, it must be emphasized that its intention is not to offer a full-fledged elaborate typology of complex nominals on the basis of their internal structure. Such a typology, with particular reference to polysynthetic languages, is developed in Mattissen (2003, 2004). Mattissen's basic distinction is between an affixal and a compositional type of morphologically complex nouns, in which the former concatenate "non-root bound morphemes" around a lexical root, while in the latter type more than one lexical root can be combined in an ad-hoc fashion. Languages of the latter structure may be subdivided into a scope-ordered and templatic type. Her typology also offers a survey of different functional and semantic categories that may be incorporated into the nominal form; however, one important difference to the present study is

that she is less concerned with the institutionalized lexical level, but rather with complexity in noun forms that may arise spontaneously in discourse (Mattissen 2003: 254-255). Furthermore, Bauer (2009) offers a typology of compounds, and Bauer (2002) investigates the semantics of derivation from a cross-linguistic perspective.

While situated in the simple present framework for the typologization of formal relations in analyzable lexical items, this section aims instead to add flesh to this rough typologization by pointing out language types that rely on one kind of complex nominal more than others and by providing examples from particular sampled languages and the language-particular word-formation devices they employ. Another aim in doing so is to highlight some of the more unusual devices for the formation of nominals that are typically not discussed in textbooks and other standard accounts of word-formation, such as nominal classifiers as well as clausal nominals as found predominantly but not exclusively in languages of North America.

4.2. A FIRST TYPOLOGY

The two very rough types of analyzable lexemes as defined in chapter 3 are those of the lexical and those of the derived type, and this classification will be the starting point for the presentation of the results. There is wide variation in the degree to which the languages of the world rely on one or the other basic type of morphologically complex lexemes. Central Yup'ik, for one, relies virtually exclusively on the derived type. On the other hand, there is a quite large number of languages in which complex lexemes are exclusively of the lexical type, as well as many languages which exhibit a mixed behavior, where both morphologically complex lexical items of the derived and lexical type are found to varying degrees. As can be seen from the histogram in figure 1, the lexical type is much more frequent on a global scale than the purely deriving type (though, as will seen below, there clearly are areal factors involved), and for those languages that fall in between the extremes, those in which items of the lexical type are more common than those of the derived type are greater in number than those with the reverse situation.

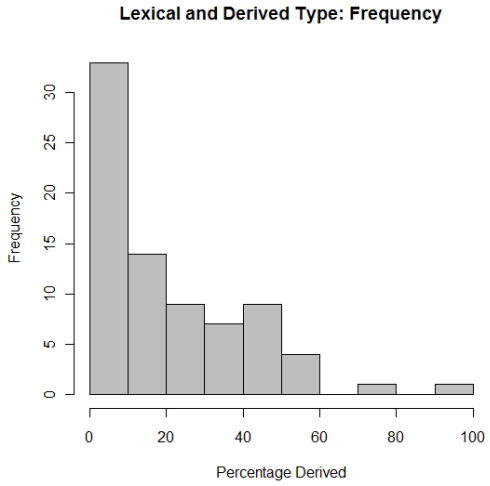


fig. 1: histogram showing the frequencies of languages with various degrees of complex items of the derived type globally

As a first approach to typologizing languages, the relative degree of derived terms in the languages of the core sample is divided into four quartiles, with languages of the first quartile (those with a preponderance for analyzable lexemes of the lexical type) constituting the first typological group, and those in the fourth quartile (those with a preponderance for analyzable lexemes of the derived type) forming the second. These two groupings define extreme types, and obviously there are many languages (those in the second and third quartile) that fall somewhere in between, that is, that have varying degrees of a mixture of analyzable items of the lexical and of the derived type. The third type of motivated lexeme which at least in some cases involves analyzability is the alternating type as defined in § 3.6.1. However, this device, compared with the lexical and the derived types, is so rare cross-linguistically that it is practically negligible for typologization (though there will be some discussion of gender alternation, the most common subtype of the alternating type in the following). This entails that the counts for the derived type and for the lexical type are practically mirror images of each other: where the derived type dominates, the lexical type is not pronounced, and vice versa.

This classification may seem a little ad hoc, but as will become clear later in the discussion, there are some noticeable typological correlates and commonalities between the languages of the extreme types.[1] Table 1 summarizes this rough typology, and provides an overview of the discussion concerning individual types that cluster at either of

[1] In general though, the ensuing discussion will be somewhat liberal given that what one is dealing with is really a continuum. Thus, languages whose percentages of derived terms are at the very top or bottom of a certain quartile are sometimes discussed along with the respective higher or lower group.

the extremes of the distribution (for percentages in the individual languages see Appendix B).

	Lexicon	Languages (in decreasing order of dominance of the derived type)	Typological profiles and reference to further discussion
Quartile I:	Derivational type dominant, lexical type subsidiary	Central Yup'ik, Burarra, Bora, Nunggubuyu, Xicotepec de Juárez Totonac, Ineseño Chumash, Nez Perce, Carib, Chukchi, Cashinahua, Chayahuita, Oneida, Buin, Rotokas, Aguaruna, Mali, Upper Chehalis, Cubeo, Aymara, Greek, Yanomámi, Efik, Khoekhoe, Khalkha	Languages with nominal classification (§ 4.4.1) Affixal type of polysynthesis (§ 4.4.2)
Quartile II + III	None of the types dominant ↕ more derived terms ↕ more lexical terms	Great Andamanese, Nuuchahnulth, Hausa, Sahu, Sora, Pipil, Arabela, Kiliwa, Laz, Kiowa, Chickasaw, Wintu, Hawaiian, Welsh, Haida, Wichí, Carrier, Kolyma Yukaghir, Kanuri, San Mateo del Mar Huave, Guaraní, Pawnee, Jarawara, Biloxi, Cavineña, Santiago Mexquititlan Otomí, Sedang, Basque, Itzaj, Bezhta, Tetun, Imbabura Quechua, Piro, Bororo, Copainalá Zoque, Kosarek Yale, Miskito, Manange, Embera, Cayapa, Rama, Bislama, Mbum, Hupda,	"Isolating" languages (§ 4.6.1) Compositional type of polysynthesis (§ 4.6.2) North American languages with verb-centered nominals (§ 4.6.4.) Mixed languages of Western Eurasia (§ 4.6.5)
Quartile IV:	Lexical type dominant, derivational type subsidiary or unattested	Abzakh Adyghe, Cheyenne, Ket, Yir Yoront, Katcha, Ngambay, Baruya, Berik, Kaluli, Kwoma, Meyah, Toaripi, Badaga, Nivkh, Kildin Saami, Highland Chontal, Wappo, Bwe Karen, White Hmong, Yay, Buli, Dongolese Nubian, Rendille, Kyaka, Mandarin, Vietnamese	"Isolating" Languages of SE Asia, the Americas, Africa and New Guinea (§ 4.5.1) Compositional Type of polysynthesis (§ 4.5.2)

table 1: relative occurrence of the derived and lexical type of complex lexemes in the languages of the core sample, divided into quartiles

It is this embryonic typology that will be the starting point for the further discussion, and as the discussion proceeds step by step through the analysis of the data here and in chapter 5, further typological correlations and dimensions will be added to the classification.

A word of caution: since this rough distinction is based exclusively on the data for the 160-meaning wordlist, the characterization is obviously restricted to the occurrence of

either formal relation in this part of the vocabulary, and need not necessarily coincide with the outcome one would obtain when considering a more inclusive set of lexical items.

4.3. AREAL DISTRIBUTION OF THE TYPES

In addition to establishing typological groupings, the areal distribution of the types is highly uneven cross-linguistically, to the effect that such a classification also leads to the emergence of areal types. A map on the basis of the core sample showing the areal distribution of the relative degree of lexical items of the derived type is seen in figure 2. The size of the dots represents the frequency of occurrence of terms of the derived type: the larger the dots, the more prevalent derived terms are among the analyzable terms (similar coding by size will be used for the visualization of other quantitative variables in subsequent chapters).

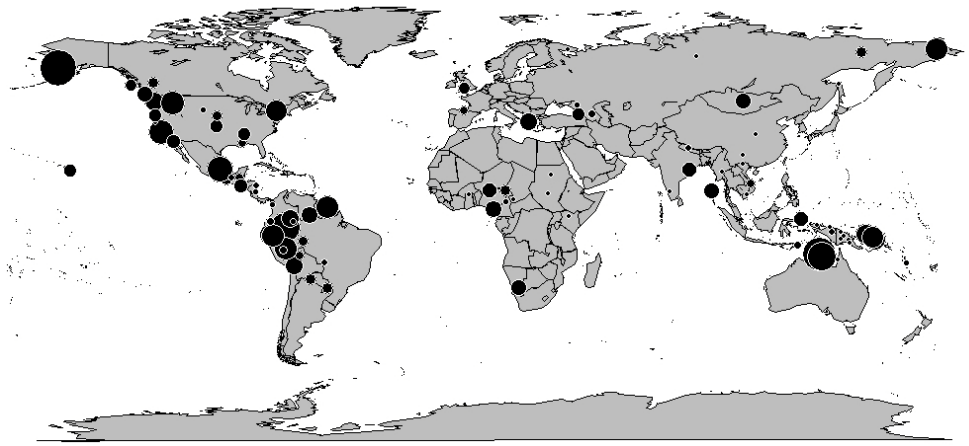


fig. 2: relative degree of lexical items of the derived type, core sample

However, it is a matter of debate and current discussion as to how areality in the distribution of linguistic features is best assessed and controlled for in typology, a problem which will be more extensively discussed when it comes to lexico-semantic patterns that appear to be areal phenomena in § 6.4.3. For the broad analysis in terms of the quantity of the main variables surveyed, as well as to avoid defining areas in a mere ad-hoc fashion that would potentially lead to the artificial emergence of areality, three different established breakdowns of the world into rather large linguistic areas are used throughout. These divisions have been defined independently of the present study to assess areality for its own sake or for quite different purposes. They are:

(i) DRYER'S (1992) SIX-AREA BREAKDOWN (Dryer-6 for short): a division of the world into six linguistic macro-areas: Africa, Eurasia, Australia-New Guinea, North America, South America, Southeast Asia and Oceania.

(ii) NICHOLS'S (1992: 25-26) ELEVEN-AREA BREAKDOWN, i.e. with Europe separated from Eurasia and Western from Eastern North America as in some of her analyses (Nichols-11 for short): Africa, New Guinea, Oceania, South and Southeast Asia, Australia, Europe, Eurasia, Western North America, Eastern North America, Mesoamerica and South America[2]

(iii) NICHOLS'S (1992: 27) THREE-WAY BREAKDOWN (Nichols-3 for short) into Old World, New World and the Pacific. The Old World comprises Africa and Eurasia (including Southeast Asia), the Pacific region is made up of New Guinea, Oceania and Australia, while the New World is constituted by the Americas

Under all three breakdowns, the areal distribution of the differential degree to which languages rely more on the derived and the lexical type is highly uneven and statistically significant (Dryer-6: $\chi^2 = 11.5158$, df = 5, $p = .04206$, Nichols-11: $\chi^2 = 23.9079$, df = 10, $p = .007849$, Nichols-3: $\chi^2 = 10.793$, df = 2, $p = .004532$, Kruskal Wallis rank sum tests). This and all further statistical analyses were carried out using the R statistical computing environment (R development core team 2009). Figures 3 to 5 plot the results for the different areal breakdowns.[3]

[2] Nichols's (1992) Ancient Near East is omitted for the obvious reason that no relevant language is present in the current sample.

[3] The plots are so-called boxplots, which are a useful way to visualize differences with respect to a variable (here: the percentage of terms of the derived type) between certain groups (here: languages of different areas of the world). The two most important types of information that can be extracted from a boxplot are: (i) the mean of the values within one group, which is represented by the thick black band in the middle of the boxes, and (ii) the variance around that mean. Specifically, the height of the boxes and the length of the whiskers (the dashed lines at both ends of the boxes) provide information as to this variance: large boxes (the size of the boxes includes 50% of the observed values) and long whiskers (for the plots shown in this thesis, whiskers extend to up to those datapoints no more than 1 time the interquartile range removed from the box) indicate high variance around the mean, while small boxes and short whiskers (or no whiskers) indicate that the observed values are grouped closely to the mean. The width of the boxes corresponds to the number of observations within each group: wide boxes show that there are many observations for that group, while relatively narrow boxes indicate that there are correspondingly fewer observations within the group. Thus, in the plots in figure 3 to 5, the fact that the boxes for the Americas are wider than those for Africa and Southeast Asia corresponds to the higher genetic diversity and hence more sampled languages for this area.

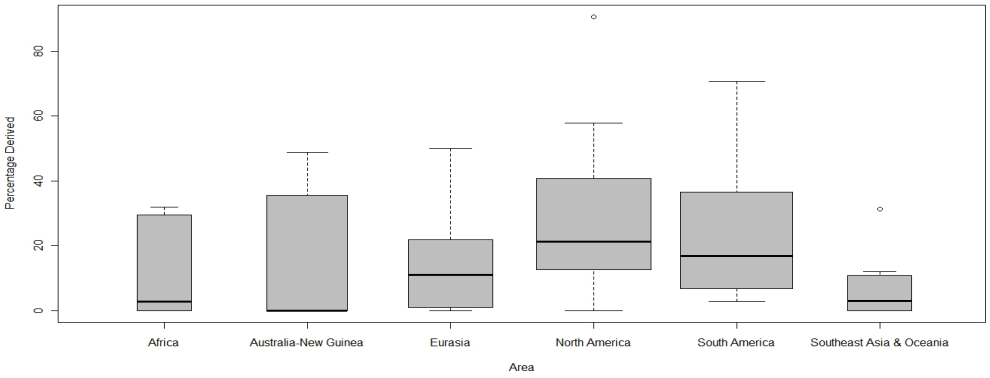


fig. 3: areal breakdown of the occurrence of terms of the derived type, using Dryer's (1992) breakdown

Here there is a lot of variation, but on average elevated degrees of derived terms in the Americas and rather low values in Africa, Australia-New Guinea, and Southeast Asia and Oceania. The Eurasian area falls somewhere in between (compare § 4.6.6.), but tends more towards low ratios of the derived type as found in Africa and Southeast Asia and Oceania. The more fine-grained Nichols-11 breakdown reveals that the distribution within North America is quite uneven, with higher degrees found in the west than in the east. Further, it reveals a difference between New Guinea and Oceania, with the former area having very few and the latter quite a lot of derived terms.

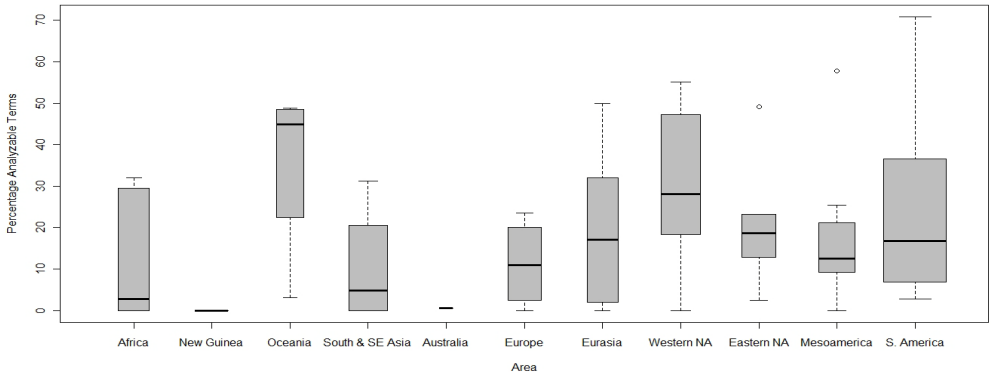


fig. 4: areal breakdown of the occurrence of terms of the derived type, using Nichols's (1992: 25-26) breakdown

Conflating these areas into the three very large areas of the Nichols-3 breakdown, one can observe a general upward trend as one moves to the New World, where the derived type is most frequent.

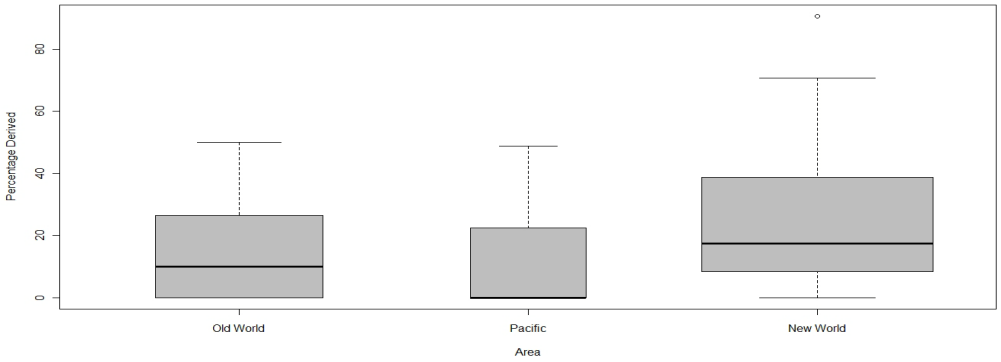


fig. 5: areal breakdown of the occurrence of terms of the derived type, using Nichols's (1992: 26-27) breakdown

4.4. DERIVED TYPE DOMINANT

4.4.1. DERIVATIONAL USE OF NOUN CLASSIFICATION SYSTEMS

4.4.1.1. Northwest Amazonia

This section is concerned with languages of South America, more exactly the Northwest Amazon area, with a profile rich in derived terms which are due to the extensive utilization of their systems of nominal classification for word-formation purposes.

In Bora and Cubeo (as well as other Tucanoan languages), there are affixal noun classifiers which occur both in the context of numerals and demonstratives, and which cannot be incorporated into the verbal complex. There is a boundary between lexical nouns and classifiers which is at times somewhat fluid; the latter may correspond partially or completely to noun roots in their segmental shape (Payne 1987: 26). This is likely due to the recent grammaticalization of classifiers from free-standing nouns (Aikhenvald 2000: ch13); the correspondence between classifier and lexical noun root is also apparent in Chayahuita, compare *-i'*, the Chayahuita classifier for liquids and the identical *i'* 'river, water.'

There is some variation in the properties of classifier systems in the Amazon discussed in Payne (1987) and Derbyshire and Payne (1990). For instance, the difference between Chayahuita and languages like Bora and Tucanoan languages is that the former allows for incorporation of classifiers into the verb complex, as seen in (1.), while the latter do not:

(1.) *a'pë-raya-t-ër-in*
burn-CL:disc-TRANS-INDICATIVE-3:SUBJ
'(he) burned (his) face'[4] (Payne 1987: 34)

[4] Glosses: TRANS 'transitive,' SUBJ 'subject.'

Furthermore, classifiers in Arabela differ from those of Huitotoan and Tucanoan languages in that they do not occur in the context of numerals (of which there are only three native ones) and accompanying demonstratives (Payne 1987: 29). Classifier systems are absent in Carib, Gê, and Tupi-Guaraní languages (although some Cariban languages may feature incipient noun classification systems, the same is true of the Tupian language Munduruku, according to Gonçalves 1987 as cited in Aikhenvald 2000). In Arabela, Chayahuita, as well as Bora and Cubeo (and other languages of their respective families), classifiers may function as nominalizers (Derbyshire and Payne 1990: 253, 266).

In fact, the first point to make is that noun classification devices provide a powerful means of word-formation that is used to enrich the lexicon. This is an aspect of nominal classification that is apparent in the set of vocabulary items under investigation here, and pertains noticeably the linguistic integration of items of acculturation, but not exclusively so.[5] However, classifiers functioning as nominalizers do not seem to be the whole story; they apparently also serve as derivational devices in general, and are not necessarily restricted to the application to verbal roots (as noted by Payne 1987: 33 for Chayahuita specifically). Noun classifiers are also present to some extent in Yanomámi. There is a class of "characteristic classifiers," which have the typical semantics expected in classifier systems: *koko* 'round/cylindrical,' *ma* 'hard,' *si* 'thin and flat,' and *mo* 'small and round' (Goodwin Gómez 2000: 15, describing a slightly divergent dialect than that of the consulted source). The origin in lexical nouns is often recognizable (such as *mo* < *mamo* 'eye, eyeball,' *si* < *si kɨ* 'skin'). Not all nouns are associated with characteristic classifiers, but most may be accompanied by so-called "quanta-specifying" classifiers. In particular, there are two such classifiers associated with plurality, *kɨ* and *pə* in Goodwin Gómez (2000) and *kɨ* and *pë* in the consulted lexical source. Semantic differences between them are hard to pin down (Goodwin Gómez 2000: 17). There is also the classifier *kɨkɨ*, apparently identical with *këkɨ* in the consulted source, glossed as 'collective' by Goodwin Gómez (2000: 16). Thus, for instance, *shi* is 'faeces' and *shi-kɨ* ~ *shi-pë* is 'guts,' *here* is 'soft,' and *here-kɨ* is 'lungs' (examples from database).

Examples from Arabela, Chayahuita, Bora and Cubeo for a selection of meanings amply illustrate a derivational use of classifiers.[6] In the slice of the lexicon investigated in the present study, derivational use of classifiers is more frequent in Bora and Cubeo, while it is found in a smaller number of examples in Arabela and Chayahuita. Perhaps this is related in some way to the typological differences in the classifier system between these languages and Bora and Cubeo as representatives of Huitotoan and Tucanoan respectively, as summarized above from Payne (1987) and Derbyshire and Payne (1990). The examples

[5] Compare also Seifart's (2005: 209fn25) statement about the Miraña dialect of Bora: "Where at all possible, Miraña favors the creation of neologisms over borrowing of morphological material. In the case of concrete objects, in particular artifacts, class markers play a major role in these creations" as well as Morse and Maxwell (1999: 73), who state that the Cubeo classifier system "enables a Cubeo who sees an unfamiliar object for the first time to give it a Cubeo name."

[6] Glosses for classifiers have been created in an ad-hoc fashion for the present study specifically, with the exception of Bora, for which Seifart's (2005) conventions are followed and for examples incorporated from sources other than the primary ones consulted.

also show that the morpho-semantic structure of the relevant lexical items is often very similar.

(2.) 'milk'
a. Arabela *quitia-aca* 'breast/teat-CLASS.LIQUID'
b. Chayahuita *sho'sho-i'* 'breast-CLASS.LIQUID'
c. Bora *mújpañejpácyo* /mújpañe-hpacyo/ 'breast-SCM.LIQUID'

(3.) 'table'
a. Bora *méétsa-wa* 'table-CL.plank'[7]
b. Cubeo *tʉoiva* /tʉoyʉ-va/
'serve.food-CLASS.BROAD.AND.FLAT.OBJECT'

(4.) 'ladder'
a. Arabela *taka-tu* 'go.up-CL'[8]
b. Bora *neríívye-wááhyo* 'ascend-CL.layered.things'
c. Cubeo *mʉi-caðava* 'ladder-CLASS.FRAMELIKE.GRIDLIKE.OBJECT'[9]

(5.) 'tear'
a. Arabela *namiji-aca* 'eye-CLASS.LIQUID'
b. Chayahuita *na'nëi'* /na'nërin-i'/ 'cry-CLASS.LIQUID'
c. Bora *máátyo-u* 'crying-CL.ROUND'
d. Cubeo *ori-yaco-rʉ* 'tear-eye-CLASS.ROUNDISH.OBJECT'

(6.) 'honey'
a. Chayahuita *nino-i'* 'beehive-CLASS.LIQUID'
b. Bora *íímúhojpácyo* /íímúho-hpácyo/ 'beehive-CL.LIQUID'
c. Cubeo *mumicoro* /mumi-córo/ 'bee/honey-CLASS.LIQUID.STATE'

(7.) 'cave'
a. Chayahuita *na'pi-të-ana* 'rock-??-CLASS.AROUND'
b. Bora *ííñuhéju* /íñu-héju/ 'earth-CL.HOLE'
c. Cubeo *cʉ̃ra-cobe* 'stone-CLASS.HOLE-LIKE.OBJECT'

[7] Note that in Bora, the word for 'table' is apparently a loan from Spanish (< *mesa*) integrated into the Bora lexicon with a suffixed classifier.
[8] From Payne (1987: 30), citing Edgar Pastor (p.c.).
[9] The Cubeo example for 'ladder' illustrates that the presence of the classifier morpheme is at times redundant or seemingly redundant on the level of the lexical item, since in this case *mʉi* alone already denotes 'ladder.'

(8.) 'nostril'
a. Arabela *naju-hua* '??-CLASS.HOLE'[10]
b. Chayahuita *nitëana* /nitë'-ana/ 'nose-CLASS.AROUND'
c. Bora *túúheju* /túú-ʔehɯ/ 'nose-SCM.HOLE'[11]
d. Cubeo *ũe-cobe* 'nose/nostril-CLASS.HOLE-LIKE.OBJECT'

While the emergence of noun classification in unrelated languages of the Amazon itself may be due to language contact (Payne 1987: 37-38, Seifart and Payne 2007: 384-385), this need not necessarily be so in the case of the morpho-semantic similarities notable in some of the above examples, although at least in some cases areal spread through language contact appears at least possible. One case is to have terms for 'tear' based on a verb root for 'cry,' which is not very common cross-linguistically (compare Appendix E, 140). Also, the fact that Bora and Chayahuita derive their terms for 'honey' from 'beehive' rather than 'bee' is not attested in other regions of the world, judging from the evidence of the sample (see Apendix E, 20).

Cubeo has more than 150 classifiers (Morse and Maxwell 1999: 75), which is one of the largest systems known, and there is evidence that they, at least in this language, also serve as singularizers, with (some of the) roots they attach to denoting masses rather than individualized entities themselves (see Lucy 1992 on this aspect of nominal classification). Thus, for instance, Cubeo *nʉo* means 'roots' generically, and *nʉo-mu* 'roots-CLASS.LIANA-LIKE' is used for a single root, *ori* denotes 'tears' collectively, and *ori-yaco-rʉ* 'tears-eye-CLASS.ROUNDISH.OBJECT' (compare 5d.) is used for a single tear. Another language spoken in the relative vicinity that has been claimed to feature a system of noun classification is Rama, of the Chibchan family. Craig (1990), however, prefers to speak of class terms, which have semantically common with classifiers that they categorize nouns depending on their shape and consistency. There are four such class terms that she accepts, three of which have a free noun counterpart: *kat* 'tree/long rigid object,' *up* 'eye, seed/round, solid object,' *kaas* 'flesh, meat/fleshy object,' and *li ~ ri*, which has no lexical counterpart and is used in association with liquids. However, Craig notes explicitly that, given that there are lexical counterparts for three items on the above list, it is difficult to distinguish between ordinary noun compounding (which is also widely used in the language involving other elements than those above) and nominal classification, and in fact does not mention any structural criteria that would allow one to do so, concluding that the phenomenon is relatively marginal in the grammar of the language.

4.4.1.2. Australia and Oceania

There are also languages in the sample outside the Northwest Amazon that employ a system of nominal classification for derivational purposes, clustering in Australia and Oceania. One of them is Burarra, spoken in Australia. In fact, the use of classifiers, which in this case come in the form of prefixes, is the prime mechanism by which the relatively few morphologically complex terms for the investigated meanings are formed in Burarra.

[10] Compare *najacua* 'nose.'
[11] Meaning of *túú* not entirely sure.

There are four noun classes, 'domestic,' marked by *mun-*, 'general,' marked by *gun-*, 'masculine,' marked by *an-*, and 'feminine,' marked by *jin-* (class labels from Glasgow 1994). Green (1987: 15) notes difficulties in figuring out a semantic basis for noun class assignment. According to Green (1987: 15), noun class is obligatorily marked on adjectives and verbs, depending on the grammatical context. In contrast, noun class is not overtly marked for most nouns (there are some nouns with a noun class prefix forming part of the stem). However, there are nouns which have variable noun classes, and these are the ones of interest here, since the semantics are altered depending on which class prefix is chosen. Examples include:

(9.) a. *mun-ngarnama* 'CLASS.DOMESTIC-inner.thigh' = 'bark'
b. *gun-ngarnama* 'CLASS.GENERAL-inner.thigh' = 'cave'
c. *gun-gapulawa* 'CLASS.GENERAL-mound.up' = 'nest'
d. *an-giya* 'CLASS.MASCULINE-egg' = 'calf'

Examples (9a.) and (9b.) not only illustrate the pervasiveness of transferring body-part terms to topological features in Australian languages (noted for instance by Schebeck 1978 and Dixon 2002: 99), but also that the classifier system provides means to derive several different meanings from the same lexical root by assigning it to different noun classes. (9c.), in contrast, illustrates an apparently purely nominalizing usage of the 'general' classifier. As for other languages of Australia in the sample, derivational use is also attested in Nunggubuyu, but appears to be much less frequent. Gurindji does not feature noun classes, and Yir Yoront is part of an area on the western side of the Cape York Peninsula where languages instead feature a productive system of compounding, involving body-part terms in particular (Dixon 1980: 111).[12]

Noun classes are also a feature of Rotokas, spoken on the island of Bougainville (see Terrill 2002 for an overview of noun classification systems in Oceania more generally). Rotokas nominal inflectional morphology is not very elaborate. There are no core cases; semantic relations otherwise susceptible to being expressed by case (e.g. locative, allative, benefactive, ablative) are coded by enclitics. Nouns are optionally inflected for number and gender, and verbs agree with the A of transitive clauses, but not with the P (see Robinson 2011: 79-92 for full discussion of Rotokas nominal morphology). Rotokas has an apparently closed set of four shape-based classifiers: *isi* 'round object,' *kuio* 'round object (edible),' *ua* 'narrow object,' and *kae* 'long object' (Robinson 2011: 50), as well as a number of classifiers which are not shape-based but instead are used for collections of certain semantic types, such as *kou* 'heap' (Robinson 2011: 51).[13] For derivational purposes, the shape-

[12] In fact, Dixon (1980: 111) notes that "many compounds would be translated by simple roots in languages from other regions" so that Yir Yoront is probably a fairly atypical language in the broader areal context of Australia, and this may be the reason for the low occurrence of the derived type in the areal evaluation above for Australia as represented by Yir Yoront. Indeed, the language has the highest percentage of morphologically complex terms of all sampled languages in Australia, although it is closely followed by Burarra. In contrast, the ratio of morphologically complex terms in Nunggubuyu, Ngaanyatjarra and Gurindji is drastically lower.

[13] Robinson (2011: 51) states that classifiers are a separate word class in Rotokas rather than being bound as suggested in earlier analyses, which is why they are written here as separate words.

based classifiers are relevant, as seen in the examples in (10.). They occur in all investigated semantic domains expect for the fourth (phases of the day and miscellanea), though they are most common in the domain of nature-related and topological terms:

(10.) a. *kaki ua* 'cracked[14] CLASS.NARROW.OBJECT' = 'cave'
b. *vuvui ua* 'transparent CLASS.NARROW.OBJECT' = 'sky'
c. *ripo kae* 'cocoa.husk CLASS.LONG.OBJECT' = 'eyelash'
d. *vovou isi* 'breathe/want CLASS.ROUND.OBJECT' = 'heart'

Some languages associated with the derived type feature other dedicated derivational morphemes in addition to noun classification devices. Here, the overall effect is thus due to a combination of the presence of noun classification employed in a derivational fashion on top of "ordinary" derivation. This is the case in Rotokas, which has the general derivational morpheme *-pa* to derive instrument and agent nouns, and exploits this resource along with its system of nominal classification in creating lexical items. It can be seen in action twice in (11.).

(11.) *ori-pa-to* *Raka* *eisi* *ruvaru-pa* *kepa*
cook-DERIV-SG.M Raka LOC heal-DERIV house
'Raka is the cook at the medical station' (Robinson 2011: 83)

The suffix *-pa* is particularly frequent in terms for items of acculturation, such as *ravireo vetaveta-pa* 'sun count-DERIV' = 'clock.'[15]

Mali, another language of Oceania, also has a system of nominal classification which is exploited for lexical expansion (Stebbins 2005: 102). Thus, the unmarked form of the root *isem* 'bird' occurs with the masculine noun class suffix *-ka* (*isem-ka* 'a bird') while, with the feminine noun class suffix *-ki*, it means 'airplane.' Similarly, compare *vilē-ka* 'crack of lightning' with *vilē-ki* 'gun,' and *slēp-ki* 'bone' with *slēp-ka* 'pen, pencil' (the feminine noun class is associated with big size in Mali according to Stebbins 2005; this is also true of many other languages, see § 6.2.3.4. for discussion). However, the classification system is also exploited to vary the semantics of native lexical morphemes. Among the examples provided are the following forms of the root *amēng* 'wood' (Stebbins 2005: 86, table 4): *amēng-ka* 'slender tree' (masculine class), *amēng-ki* 'large full grown tree' (feminine class), *amēng-ini* 'stick' (diminutive class), *amēng-ēm* 'tree stump' (reduced class), *amēng-igl* 'plank' (excised class), *amēng-vēt* 'pole' (long class), and *amēng-ia* 'large log' (extended class).

A system of nominal classification or noun classes might also be a feature of Great Andamanese, although the situation remains unclear on the basis of the available evidence (also note that Great Andamanese is in the second quartile of the division made above, but on the top edge). Manoharan (1989: 61) says that Great Andamanese "uses a stem or root

[14] Full gloss: 'opened, cracked, split open.'
[15] At times, gender/number suffixes appear to be employed in a derivational fashion as well, such as in *kaporo-to* 'space-M.SG' 'scissors,' but this type of formation remains unclear.

for one concept" and then may vary the semantics of that concept "differently using different affixes what [sic!] are called 'formative affixes'." Examples include (adapted from Manoharan's 1989: 61-63 more extensive list):

(12.) a. *-beːc:*
i-beːc 'honey comb'
araː-beːc 'tail (of birds)'
tot-beːc 'feather'
ṭɔt-beːc 'hair-my'
ṭɔw-ter-beːc 'cloud' (*ṭɔːw* 'sky')
ṅeruluto-beːc 'eyelash-your' (*ṅe-erulu* 'eye-your')
ṅertaːp-beːc 'beard-your' (*ṅe-ertaːp* 'your-chin')[16]
taːjiyo-tot beːc 'bird' (*ta jiyo* 'pool')

b. *ɸoṅ* 'hole, cave'
myoː-ɸoṅ 'cave' (*myoː* 'rock')
inɔ-tɛr-ɸoṅ 'well' (*inɔ-* 'water')
ṅerkɔːtho taraː-ɸoṅ 'nostril-your' (*ṅerkɔːtho* 'nose-your')
ṅa-ɸoṅ 'mouth-your' (*ṅa-* 'your')[17]
ñɔm-ɸoṅ 'armpit'

Burenhult (1996: 10) calls these "formative prefixes," but from the examples in Manoharan (1989: 61-64) it seems that the recurrent elements in fact more often, but not exclusively, follow rather than precede other material. Given the variability in position extractable from the examples, it is unclear whether one is dealing here with something other than "ordinary" compounding. A system of body part classification that is extended to the domain of kinship is mentioned by Abbi (2006b). Since the majority of examples do not in fact involve body-parts as the putatively classified entity, it is unclear how the system of these affixes (if it is a system) is related to the system of body-part classification outlined in Abbi (2006b); Manoharan (1986: 28) speaks of two different "types," which suggests that they are independent of one another. There are also similar prefixes in semantically related terms: compare for instance *âkà-bang* 'mouth' with *âkà-pai* and *âkà-pē* 'lip' and *âkà-êkib* 'jaw,' *ông-pâg* 'foot' with *ông-tôgur* 'ankle' etc. To add to the confusion, these prefixes are described as indicating the possessor of body parts in Man (1923), although they clearly also occur on terms not denoting body-parts. Further, the shape of the prefixes does not match those mentioned by Abbi (2006), although again it may be relevant that present-day Great Andamanese is a conglomerate of several erstwhile distinct

[16] Compare Abbi (2006a: 31): *ʈhɛr-ʈap'-bec'* 'my beard'

[17] But compare Abbi (2006a: 31) *tʰɛr-ɸoŋ* 'my mouth.' *ʈʰa*, which should correspond orthographically to *ṅa-* in Manoharan is used for 'tongue' and kinship terminology according to her, though note that there is variation between speakers and that "[c]onsistency is lacking, as the language contains several varieties of past and present dialects. In general, a speaker has two or more forms in the verbal repertoire that s/he can vary freely in all contexts ..." (Abbi 2006b: 114).

varieties or languages (Manoharan 1986: 27-29 reports a wide range of allomorphy in present-day Andamanese resulting from this mixture, and provides the following list of attested prefixes: *it-*, *et-*, *ir-*, *er-*, *akka: ~ ekka:-*, *ara: ~ era:-*, *atta: ~ etta:-*, *ot-*, *tot-*, *ta:-*, *tara:-*, *i-*, *0-*). The discussion in Manoharan (1986: 28) suggests an extremely interesting system of classification based on conceptual dependency. As his example goes, 'pig' is independent, but 'head' is not in that it is relational and attached to a larger entity, and thus the latter receives marking with a "classifier." Compare *mōlo* 'egg' from the database with *âr-mōl* 'bud' (this association is also found elsewhere, see Appendix E, 9), which would fit nicely into this scheme: an 'egg' is conceptually an independent unit, since it can be picked up and carried away (it is detached, in terms of Gibson 1979), while a 'bud' is attached to a larger structure, namely the budding plant, and as such does not enjoy the same degree of conceptual independence and is marked by a formative prefix attached to the same lexical root. Manoharan (1986: 30) also emphasizes that dialect mixture may be responsible for blurring an erstwhile more straightforwardly semantically based system of formative prefixes assignment, with vestigial traces of the putative original system left. For instance, the marker *ara: ~ era:-* is found in many names of objects at the back of the larger structure (e.g. *ara:-be:c* 'tail of bird,' *ara:-ce* 'sting,' *era:-bu:-co* 'roots'); note also the transition in Bantu from a putatively initially semantically based to a formally based system of assignment (Corbett 1991: 48-49). Manoharan (1986: 30-31) also attaches importance to the appraisal that the Andamanese system is different from the classifiers found in languages of mainland India: "they are not here called classifiers since the primary function of these formative affixes is to change the meaning of the concept attached to the root and not to classify the nouns or that of the kind" and summarizes: "in the case of formative affixes, the language does not possess different words for different concepts; rather, it changes the meaning of the primary concept into the specific meaning." Regardless of the precise properties of the Andamanese system which is not entirely clear on the basis of the available evidence, this is precisely the aspect which is relevant for the present discussion.[18]

4.4.1.3. Classifiers in Southeast Asia

Classifiers are also present in Southeast Asian languages.[19] In line with the isolating typological profiles of these languages, they are typically independent words in these languages rather than affixal, as is predominantly the case in the Amazon region (see Aikhenvald 2007: 10). Burmese data from Becker (1975) presented in Aikhenvald (2007: 11) indicate that the numeral classifiers in Southeast Asian languages may be used to variegate the lexical semantics of nouns in discourse and to highlight aspects of their meaning (see also Bisang 1999: 129-130). However, classifiers in Southeast Asian languages appear to play a much smaller role as derivational devices and thus as a means to enrich the lexicon when compared with the situation in the Amazon; Bisang (1999) surveys functions of classifiers in a number of Southeast Asian languages, but does not mention derivational

[18] Compare Burenhult (1996: 10): "[i]t is extremely difficult to get a clear idea of this system of formative prefixes, and the lack of raw data prevents us from taking the analysis any further..."

[19] There are also entities called "verbal classifiers" in languages of North America, for instance Haida, which serve to specify the semantics of the verb they attach to; these elements fall outside the scope of the present discussion because of being associated with verbs without changing word-class.

usage. However, Bisang (1999: 114) mentions that the "high degree of indeterminateness of nouns is a very important characteristic of East and Southeast Asian languages ... which is crucial to the existence of classifiers." Still, this indeterminateness does not appear to be resolved at the level of the lexical item by using classifiers in a derivational fashion, but rather to modulate the semantics of the classified noun at the discourse-level (though see also § 4.5.1.3.2. for some evidence from Yay). Consequently, most Southeast Asian languages in the present sample are found in the fourth quartile, with terms of the derived type exceedingly rare (compare discussion in § 4.5.1.3.)

Moving back to the general discussion, it has to be pointed out that the sheer presence of a system of nominal classification does not necessarily entail its employment for derivational purposes. Nominal classification is also found in Meyah (Gravelle 2004) and yet, judging from the database, it does not appear to make use of its noun classification system on a large scale for word-formation. Here, differences in the individual system, such as whether classification occurs predominantly in the context of numerals, obviously may play a role.

4.4.1.4 Excursus: Noun Classes and Gender

Noun classes are also found in languages of Africa, most notably in languages of the Bantu subgroup of the Niger-Congo family, and here, the relevant markers are also employed in a derivational fashion. In terms of Corbett (1991, 2007), genders are identified on the basis of syntactic evidence, or, more precisely, by the ability of the gender of a head noun to trigger agreement on other elements of the sentence and thus to constitute agreement classes. In Bantu languages such as Swahili, gender is indicated by a prefix on noun roots. A complicating aspect is that prefixes on nouns do not always match the agreement forms (Corbett 1991: 44-45), a fact that makes it necessary to distinguish between agreement classes, which are relevant evidence for identifying gender, and inflectional classes of noun roots. It is the latter class which is relevant here, since indeed these inflectional classes may be used in a derivational fashion.

In Swahili, for instance, markers for inflectional class may be prefixed to a nominal or verbal root to form the names for artifacts (13a.), and in the domain of natural phenomena and body parts, the same root may be used with varying inflectional prefixes to alter and specify the semantics encoded in the root (see also Corbett 1991: 44), as in (13b.)

(13.) a. *u-funguo* 'key' (*fungua* 'to open')
b. *u-tumbo* 'guts' (*tumbo* 'belly, stomach')

Mufwene (1980) emphasizes that class prefixes of the Bantu type are not just inflectional, but also derivational, arguing on the basis of similar evidence to that presented here. Swahili, however, does not exclusively rely on noun class alternation as a motivating device, but frequently employs syntactic devices as well to create new lexemes in phrasal form, for instance *shavu la mguu* 'cheek of leg' = 'calf.'

Having moved the discussion to Bantu-type noun class systems, a review of gender is not far away. In fact, noun class systems are frequently subsumed under the notion of gender, for instance by Corbett (1991), who distinguishes between semantic and formal

systems of gender assignment and discusses noun classes in Bantu languages as being of the formal kind (although they have a semantic component and may have started out diachronically as being assigned on a semantic basis alone, Corbett 1991: 48-49). Furthermore, gender systems are sometimes employed to alter the semantics of lexical items by assigning the same lexical item to different genders, for instance, German *die See* (feminine) 'sea,' *der See* (masculine) 'lake.' Scattered examples of this phenomenon are found in a number of languages in the sample. However, one language that frequently utilizes variation in the assignment to genders of the same lexical items as a motivational device is the Choco language Embera, as seen in the examples in (14.) (glossing is somewhat simplified here for ease of presentation, and some of the stated meanings in the source are omitted).

(14.) a. *hú* 'breast, chest' (masc), 'teat, nipple' (fem.)
b. *kẽŋgú* 'nerve' (masc.), 'vein, lode' (fem.)
c. *kathárro* 'catarrh' (masc.), 'phlegm' (neut.), 'snot' (fem.)
d. *káarta* 'paper' (masc.), 'letter' (fem.)

As examples (14c.) and (14d.) show, variability in gender assignment is also present in loanwords from Spanish.

Moving to a general discussion of noun classification from the perspective of lexical typology specifically, an important aspect of derivation by noun classification is noted by Aikhenvald (2000: 267): "Noun classifiers are often used to highlight different aspects of the meaning of a polysemous noun." This can be nicely illustrated with examples from Cubeo. For instance, the root *cu̱ra* in (6c.) only occurs suffixed with classifiers to the effect that its semantics can in fact only be extracted by comparing that of the different terms with classifier and "subtracting" their respective meaning. As another example from the consulted source, the root *ũe* in fact only occurs suffixed with classifiers. With *-ca* 'tuber-shaped,' the derived form means 'nose, snout,' with *-bo* 'round or hard,' it means 'nose of an airplane,' and with *-cobe*, as in the above example, 'nostril.' Thus, "[t]he ways the presence of classifiers in a language can reduce the actual lexicon by expressing the same concepts through grammar relates to a well-known facet of the interface between grammar and lexicon" (Aikhenvald 2000: 268). [20] From the point of view of lexical typology, another property of the languages with noun classification devices in the sample, in particular those of the Northwest Amazon, is also noteworthy, namely that they lack patterns of colexification that are frequent elsewhere. This is particularly noticeable for the meaning pairs 'milk' and 'breast' (in all the languages surveyed above 'milk' is derived from 'breast' by means of a classifier) as well as the lack of colexification of 'bee' and 'honey,' which is a less frequent, but still common enough phenomenon cross-linguistically (see Appendix E, 30 and 123). It seems that the presence of a classifier system leads these languages to favor expressing these meaning pairs by derivation rather than colexification. This would need to be assessed against a larger number of classifier languages of the

[20] As Frank Seifart (p.c.) points out, this can only be said to the extent that classifiers actually are grammar. As noted in the above discussion, many classifiers can be traced back diachronically to free-standing nominal roots, and in some cases, they are in fact identical in shape to nouns.

Northwest Amazon, but there does appear to be a (weak) correlation, although probably no strict law, to this effect.

4.4.2. AFFIXAL TYPE OF POLYSYNTHESIS

This type is mostly found in the American Northwest, and is chiefly responsible for the high frequency in derived terms there. It is most clearly represented by Central Yup'ik, an Eskimo-Aleut language spoken in Alaska. Like other Eskimo-Aleut languages, it is notable for the absence of compounding, employing only derivational mechanisms for word-formation purposes (Mather et al. 2002: 16, although Mithun 2009: 13 argues that noun-suffix structures in Eskimo-Aleut languages have recently originated from incorporation-like structures). Derivation in Central Yup'ik operates by a class of morphemes called postbases in Eskimo-Aleut studies. These are exclusively suffixal, and may be applied iteratively, that is, a lexical root (and note that the defining feature of this type in the present study is that only one lexical root be involved!) may be suffixed with several postbases that add to or modify their meaning. According to Mather et al. (2002: 33), postbases in Central Yup'ik can be classified according to whether they are trans-categorical or not, that is, whether they change the lexical class of the root (or of the root already modified by postbases, which is called an expanded base) they attach to (verbalizing/nominalizing postbases) or whether they preserve the lexical class of the root (or expanded base) they attach to (nominal-elaborating postbases/verb-elaborating postbases).

The following construction (from Mather et al. 2002: 35, glosses inferred) illustrates both types:

(15.) *teng-ssuu-ksuar-Ø*
fly-instrument-small-ABS
'small plane (flying instrument)'

Teng is a verbal root meaning 'to fly,' *-ssuu* is a nominalizing postbase that creates instrument nouns, and the following *-ksuar* 'small' is a nominal-elaborating postbase adding the meaning 'small' to the preceding material. Finally, there is a zero-marked absolutive case marker (nominals are obligatorily inflected for case). As the example also shows, postbases may be used in an ad-hoc fashion to modify the semantics of a base or expanded base, but they are also the prime mechanism responsible for the formation of institutionalized morphologically complex expressions in all semantic domains under investigation here (Fortescue 1980: 259, in discussing the related language West Greenlandic, reports frequent conventionalization of combinations of root and postbase and notes in general that in polysynthetic languages typically "the area of overlap between clearly lexicalized items and productive rules which can generate them from more basic morphemes results in an extensive descriptive 'penumbra.'"). Examples from the database (some of the terms only occur in some dialects of Central Yup'ik) include:

(16.) a. *yaqulek* /yaquk-lek/
'wing-one.having' = 'bird, duck, fowl, angel'
b. *merr'aq* /meq-rraq/ 'fresh.water-a.little.bit.of' = 'dew'
c. *agluryak* /agluq-yak/
'center.beam.of.a.structure-thing.similar.to' = 'rainbow'
d. *mingqun* /mingqe-(u)n/ 'sew-device.for' = 'needle'
e. *uquryak* /uquq-yak/ 'oil-thing.similar.to' = 'heavy sweat'

Note that by means of the postbase *-yak* 'thing similar to' the grammar of Central Yup'ik has a "built-in mechanism" to form metaphor-driven denominations. Still, contiguity dominates as the device underlying the motivated lexical items of Central Yup'ik. Note also that, in spite of the lack of means to combine lexical roots in the literature, there are some examples in the database that involve the combination or at least loose juxtaposition of two lexical roots such as *agyam anaa* 'star faeces' = 'meteor;' however, these are clearly marginal.

Languages which have a relatively high degree of analyzable lexical items of the derived type are common in the Americas, among them Nuuchahnulth (lacking both compounding and incorporation according to Davidson 2002: 92, although there are a number of examples of the database in which *mis* 'thing,' which has the potential to occur as a free-standing form, acts as a nominalizer), Nez Perce, Upper Chehalis, Ineseño Chumash in the Northwest and Xicotepec de Juárez Totonac in Mesoamerica. Some examples from Ineseño Chumash are in (17.).

(17.) a. *š-oqyokonič*
'3SG/3SG.POSS-be.uneven/protruding' = 'Adam's Apple'
b. *is-xɨp* 'one's.own-stone' = 'testicle'
c. *'aqmilimu'* /'aqmil-mu'/ 'drink-DERIV' = 'spring, drinking place'
d. *s-axiyi'* '3SG/3SG.POSS-be.dark' = 'night'

Derivation also plays a major role in word-formation in Chukchi, a polysynthetic language of Eurasia (see Dunn 1999: 143-148 for an overview of derivational devices). Complex lexemes of the derived type include:

(18.) a. *təlan* /təle-n/ 'go-LOC' = 'path' (see also Dunn 1999: 146)
b. *ajmə-n* 'fetch.water-LOC' = 'hole in ice, well'
c. *ine-newntet-icyən* 'ANTIPASS-open-INSTR' = 'key, opener'[21]

However, Chukchi also allows for the concatenation of a variety of form classes into a nominal word form (see Mattissen 2003: 255-256 and Spencer 1995) and is consequently

[21] Verb stems must be intransitive for the instrumental suffix to be applicable; transitive verbs therefore must undergo antipassivization (or incorporation) first (Dunn 1999: 147); also note that the shape of the suffix is given as *-ineŋ(e-)* by Dunn (1999:147).

typologized as belonging to the compositional type of polysynthetic languages by Mattissen. Indeed, formations of the derived type are relatively rare in Chukchi when compared with Central Yup'ik or Nuuchahnulth (see brief discussion in 4.6.4.3.4.), although still more frequent than in most languages in the sample.

4.5. LEXICAL TYPE DOMINANT

4.5.1. ISOLATING TYPE

There is a discernible lexical profile of languages with predominantly isolating structure which involves analyzable terms of the lexical type as defined in § 3.6.1. (compare Aikhenvald 2007: 24, Vajda 2004: 400 and more tentatively Bauer 2009: 355 on the affinity of isolating languages with compounding). What it means to be an "isolating" language is, however, somewhat ill-defined. Following Greenberg (1960: 186), a relatively low ratio of inflectional morphemes per word may be taken as a hallmark of a prototypically isolating language; this is assessed here impressionistically for the time being, with reference to bound morphology in general, that is, inflectional morphology is frequently included in the discussions. Although inflectional morphology is not directly relevant for lexicalization processes and word-formation per se, the goal here is to show that the profile pertains to the morphological typology of the languages at large. This impressionistic assessment is to be systematized and operationalized in a bit more detail later. The isolating profile is illustrated with examples from languages from four different regions of the world: Africa, New Guinea, Southeast Asia, and the Americas.

4.5.1.1. Africa

4.5.1.1.1. *Ngambay.* Ngambay, a Nilo-Saharan language spoken in the Central African Republic, is a language in which all analyzable lexemes in the database are of the lexical type (see Moser 2004: 130 for the importance of compounding specifically). While Ngambay clearly features bound morphology, it is relatively sparse. For instance, there is no case marking for core grammatical relations, although there are inflectional categories such as inalienable possession as well as other non-bound case-like forms expressing relations such as dative and associative. Plurality is expressed by an independent word. Ngambay verbs may show agreement with pronominal subjects under certain conditions (Moser 2004: 290), and there is also a set of object suffixes (Moser 2004: 94). Analyzable terms in Ngambay range from the most common combination of two elements to phrasal idioms (still called compounds by Moser 2004: 132).

(19.) a. *ddéw-màann* 'way-water' = 'river'
b. *nɔn-kake* 'tear-tree' = 'resin'
c. *kàrè wùr énje* 'sun heart mother' = 'noon'
d. *bɔ̀-dìngàw wòy dɔ̀ ji-á té* 'big-man dies head arm-3S LOC' = 'sunset'
(Moser 2004: 132, small caps added)

Ngambay also features complex lexical items with *né* 'thing' such as *né kìsì* 'thing sit.down' = 'chair.' Moser (2004: 133) regards these as deverbal nominalization, and indeed, semantically they are equivalent to plain nominalization in languages with dedicated nominalizers. Since, however, *né* can also occur as a noun on its own, as shown by the example in (20.), such Ngambay nouns are considered to be of the lexical type here.

(20.) *gate né nèénn bàánn wà?*
price thing DEM how much QUE
'The price of this thing (is) how much?'[22] (Moser 2004: 300)

4.5.1.1.2. *Mbum.* Mbum, of the Adamawa branch of Niger-Congo, is a fairly isolating language, that is, the ratio of bound morphemes to free forms is low. The example in (21.), inferred from Hagege (1970: 209), illustrates some typical properties of this language type, such as lack of agreement on the verb and expression of plurality by an independent word rather than an affix.

(21.) *nzùk kà híŋà pínà Bí pàì rí*
man/human ASPECT do work in field PL
'the man/person is working in the fields'

Mbum has word-class changing derivation (Hagege 1970: 128-155), including processes to derive nominals. However, judging from Hagege's examples, the output of these processes are mostly abstract nouns, while the creation of terms for objects in the extra-linguistic world is largely taken over by complex lexemes of the lexical type (there is also one term in the data for the present study tentatively classified as being of the derived type, namely *ɓì-mbàm* 'at-rain' = 'puddle').

In the Mbum lexicon, there are basically two types of morphologically complex lexical items, both of the lexical type. The first involves the direct juxtaposition of two lexical roots which may be of different parts of speech (see Hagege 1970: 159-188 for more detailed discussion). Hagege calls the resulting structures compounds ("composes"), but gives no criteria concerning how to distinguish them from syntactic phrases. Some examples from the database are in (22.).

(22.) a. *ŋgàŋ-kpù* 'skin-trunk/tree' = 'bark'
b. *nûm-jóró* 'fat-bee' = 'honey'
c. *fè-nâm* 'thing-sleep' = 'bed, lit'

The second structural type, which is somewhat rarer in the database, consists of complex lexical items constructed using the connective morpheme *à* which is also found in Mbum possessive constructions (this marker is glossed as "juncteur" in Hagege 1970 and "connective" in Hagège 1993; the latter convention is used here). Examples are in (23.).

[22] Additional gloss: QUE 'Question.'

(23.) a. *pàk à jáù* 'house CONN urine' = 'bladder'
b. *pàk à ndòì* 'house CONN bird' = 'nest'
c. *sàk à mbàm* 'tear CONN rain' = 'lightning'

Hagege (1970: 161-162) also mentions more complex lexemes, some of which have phrasal structure, such as *pélé mì á nzì mbáp* 'tomorrow I ASPECT be.big too' = 'pinky.' Metaphor-driven conceptualizations in complex terms dominate in all domains, with the exception of artifact terms. In the artifact domain, like in (22c.), compounds with *fè* 'thing' abound. In fact, compounds with *fè* 'thing' and *nzùk* 'man, human' are the common way in Mbum to create agent and instrument nouns (Hagege 1970: 186), a function taken over by dedicated derivational morphemes in other languages.

4.5.1.2. Americas

4.5.1.2.1. *Miskito.* Languages with a similar lexical profile are also found in the Americas, although they are relatively rare in this area. One case in point is Miskito of the small Misumalpan family spoken in Nicaragua. This language features very little inflectional nominal morphology apart from possessive suffixes; the verb agrees with the subject and under certain conditions with the object in number, and the subject marking paradigm is intertwined with tense (Hale 1994: 264), as seen in (24.).

(24.) *Witin raks wal sula kum ik-an*
He gun with deer one kill-PAST3
'he killed a deer with the gun' (Hale 1994: 264)

In this language, morphologically complex items are exclusively of the lexical type; one such for an item of acculturation is seen in context in example (25.).

(25.) *Yang dusa pihni di-aia want sna*
I stick white smoke-INF want-PRES1[23]
'I want to smoke a cigarette.' (Hale 1994: 270)

Complex items in Miskito may involve verbal, nominal, and adjectival constituents (Conzemius 1929: 75 notes that "[m]any nouns, adjectives and verbs of the languages under consideration have the same root," which suggests that the distinction between noun and verb may be weak on the lexical level), as seen in examples in (26.).

(26.) a. *il byara* 'water abdomen' = 'valley'
b. *pauta yuya* 'fire grain' = 'spark'
c. *mita sirpi(ka)* 'hand little' = 'finger'
d. *ki mita* 'lock hand' = 'key'

[23] Additional gloss: PRES1 'present tense, 1st person.'

Terms for artifacts are either complex formations, as (26d.), or are loanwords from English, sometimes via Miskito Coast Creole English. Next to terms of the lexical type, Miskito also features derived terms formed with the suffix *-ka*, such as *aiklabaika* /aiklab-aia-ka/ 'fight-INF-DERIV' = 'weapon,' but they are clearly outnumbered by analyzable terms of the lexical type.

4.5.1.2.2. *Wappo.* The lexical type of complex lexemes is also dominant in Wappo, formerly spoken in California, and with this lexical profile the language is in contrast to many languages of North America. It would be incorrect to characterize Wappo as an isolating language since there is an apparatus of bound morphology associated with both verbs and nouns. As described by Thompson et al. (2005), nouns inflect for a number of cases, including the core cases nominative and accusative, while verbs are inflected by suffixes for tense/aspect and may carry directional prefixes. Mood is specified by preverbal particles. Wappo inflectional morphology is however comparably limited, lacking for instance the typical ingredients of polysynthesis such as incorporation and pronominal affixes cross-referencing information of NPs on the verb.[24] Pronouns are independent words rather than affixes, and there does not seem to be any overt agreement marking on the verb with its arguments. Furthermore, judging from the examples in Thompson et al. (2005), inflectional synthesis is rather low in actual usage, with inflection on nouns being mostly restricted to case, and most verbs in the examples bearing one or at most two inflectional morphemes. Wappo is included in this discussion, since its inflectional possibilities, while not non-existent, are limited when put in an areal-typological perspective, and there is a correspondingly higher degree of complex lexemes of the lexical type that appears to correlate with this situation.

Thompson et al. (2005: 9) mention associative phrases, consisting of two juxtaposed nouns (27.):

(27.) *oye? šukolo?-i* *pico:we-khi?*
pot bottom-NOM dirty-STAT[25]
'the bottom of the pot is dirty' (Thompson et al. 2005: 9)

This construction is also used to indicate inalienable possession, leading these construction to be indistinguishable from associative phrases (Thompson et al. 2005: 15), as seen in (28.).

(28.) *ċiċa khap-i* *ke?te-khi?*
bird wing-NOM broken-STAT
'The bird's wing is broken' (Thompson et al. 2005: 15)

[24] No rigorous attempt is made here to define the notion of "polysynthetic language" (see e.g. Mattissen 2003, Mithun 2009 for definitional issues).

[25] Additional glosses: NOM 'nominative case'

At any rate, combinations of two nouns, whether at the syntactic or at the morphological level, make up the largest proportion of analyzable lexical items in Wappo, although sometimes more complex structures are found (29c.) and the mechanism may be applied recursively (compare 29d. with 29e.):

(29.) a. *holchíla* /hól-chíla/ 'tree-skin' = 'bark, thin bark'
b. *huy-nán* 'breast-mouth' = 'nipple, teat'
c. *lékiš ċiti·-wélma* 'swallow bone-protection' = 'Adam's Apple'
d. *hu-méy* 'head-water' = 'tear'
e. *tuš-huméy* 'bee-tear' = 'honey'

4.5.1.3. Southeast Asia

4.5.1.3.1. *Vietnamese.* Languages of Southeast Asia are known for their isolating character, and indeed, the language used here to illustrate the prevalence of the lexical type of analyzable terms said to be associated with isolating languages, Vietnamese, is typically used in textbooks to illustrate a prototypically isolating language. However, obviously, in spite of the alleged one-to-one match between word and morpheme, Vietnamese is not devoid of analyzable items, particularly compounds. Thompson (1991: 126-138) provides an overview of different subtypes of compounds in Vietnamese, noting that in general it is quite difficult to distinguish them from syntactic phrases. The basic distinction is one in so-called syntactic and non-syntactic compounds (next to so-called "idiom compounds"): for the former, there are parallel syntactic phrases containing the same elements in the same order (one criterion available to distinguish them is differences in stress) and non-syntactic compounds, for which there is no parallel syntactic construction. Among syntactic compounds, there are those of the dvandva or coordinating type (called "generalizing compounds" by Thompson 1991: 128), such as *bàn-ghế* 'table-chair' = 'furniture' (paralleled syntactically by *bàn ghế* 'tables and chairs'), endocentric ("specializing") compounds, such as *dây thép* 'string/cord/wire-steel' (paralleled syntactically by *dây thép* 'electric wire). Among the nonsyntactic compounds is a type in which synonyms are concatenated (similar to Mandarin Chinese) called "reinforcing compounds;" of this type, only verbal compounds are given as examples by Thompson (1991: 130-131). Furthermore, among the nonsyntactic compounds there are endocentric compounds (called "attributive compounds" by Thompson 1991: 132), for instance *học trò* 'study school.age.child' = 'schoolchild, pupil'). According to Thompson, many examples of this type of compound are partially composed of loans from Chinese, and there are also some which are even calqued from Chinese. Chinese influence is also responsible for a class of compounds called "pseudo-compounds" by Thompson (1991: 133). This type involves bound morphemes (as defined by Thompson 1991: 118, that is, morphemes that cannot appear as a free form or in combination with a construction larger than a basic free form), some, but not all of which are borrowings from Chinese. Semantic relations in this type of compound are similar to those in non-pseudo-compounds (see Thompson 1991: 134-136).

As Thompson (1991: 136) notes, most of the compounds and pseudo-compounds are disyllabic (and hence, bimorphemic). Occasionally, however, compounding may apply recursively to yield words of three or four morphemes, such as *vô-tuyến điện-thoại* 'with-

out/lacking-wire/line electricity-speech/conversation' ='radio telephone.' However, a dispreference for such longer forms is noted, and shortening of three-syllable compounds to two syllables is observed (Thompson 1991: 137). This is in line with the data in the database. Combinations of two morphemes are clearly dominant in all semantic domains, although some words consisting of three elements can be found. Some examples are below (note that 30c. is trisyllabic/trimorphemic):

(30.) a. *lông chim* 'body.hair bird' = 'feather'
b. *sương mù* 'dew blind' = 'fog'
c. *dòng nước xoáy* 'current water swirl' = 'whirlpool'
d. *xe lửa* 'vehicle fire' = 'train'
e. *cổ chân* 'neck leg' = 'ankle'

There is also reduplication of various types (discussed under the heading "derivatives" by Thompson 1991: 139-140). Reduplication results in an iterative, augmentative, or other meaning, as in *nói nói* 'keep talking and talking' (*nói* 'talk').

4.5.1.3.2. *Yay.* Yay is another Southeast Asian language with the typical ingredients of the isolating profile. Grammatical relations are signalled not by affixes on either noun or verb, but by word order which is typically SVO:

(31.) *ku*[1] *mi*[4] *θoŋ*[1] *tua*[4] *ya*[5]
I have two CLF wive
'I have two wives' (Hudak 1991: xxvi, small caps added)

Example (31.) also demonstrates the absence of number marking on nouns, as well as the presence of a classifier system for most count nouns (Hudak 1991: xxvii). Verbs do not inflect for tense or number, and causatives are periphrastic:

(32.) *van*[4]*van*[4] *hauɥ*[3] *kwaat*[2] *hay*[6] *vaay*[4]
every.day cause sweep.out dung buffalo
'Every day he has him sweep out the buffalo dung'
(Hudak 1991: xxix, glossing slightly modified)

Another feature typical of isolating languages (though not exclusively found here) are serial verbs that may be thought of as making up for the absence of morphological means to express certain grammatical categories (Aikhenvald 2006: 53); Yay has them, as seen in (33.).

(33.) *may*[6] *faay*[4] *koŋ*[2] *ma*[1] *rup*[1] *caw*[3] *hauɥ*[3] *ku*[1]
bamboo bend come stroke head give I
'The bamboo bends down to stroke my head for me' (Hudak 1991: xxx)

The majority of analyzable items are made up of two elements with both nominal and verbal meanings. Since possession is expressed by the simple juxtaposition of possessor and possessed, such as *paw²po⁵ ti⁵* 'father he' = 'his father' (Hudak 1991: xxvi), there is no immediately obvious structural difference between these and possessive constructions, like in Wappo. Occasionally, however, concatenation of three elements is found.

(34.) a. *ram⁶ foŋ⁴* 'water roof' = 'wave'
b. *rua⁴ bin¹* 'boat fly' = 'airplane'
c. *ʔaay¹ ka¹ (raay⁴)* 'goiter leg (mark)' = 'calf'
d. *raan⁴ pay¹ rok⁵* 'house go outside' = 'outdoor toilet'

Quite often, analyzable terms in Yay are semantically redundant. This may be due to either one member of the complex lexical item already having the same semantic extension as the whole complex term, or the presence of a classifying morpheme.

(35.) a. *θɯay1 (ram⁶)* 'steam (water)' = 'steam'
b. *(θan³) rɯaŋ¹* '(CLASS.ROPE) tail' = 'tail'

Occasionally, there appear to be borderline cases in two senses: first, it is sometimes unclear whether a given morpheme is employed in a classifier-like fashion, and secondly, there may also be some differences in semantic nuances between simplex and complex terms. For instance, *ta⁵* 'river' may be combined with *ka¹* 'leg' to yield the meaning 'a river, a very big river.' If indeed *ka¹* 'leg' serves as a kind of classifier (which is suggested by the parallel construction *ka¹ ran¹* 'leg road' = 'road'), then the semantic difference between *ta⁵* and *ka¹ ta⁵* may be an instance of derivational usage of classifiers in Southeast Asia.

However this may be, it is worth pointing out that not all languages of Southeast Asia are exclusively of the lexical type. In another language, Sedang (Austro-Asiatic, Mon-Khmer), there is also a derivational apparatus exploitable for word-formation, including the nominalizing prefix *kơ-* and the nominalizing infix <*ơn*>. This feature is one of the main characteristics distinguishing many Mon-Khmer languages from other Mainland Southeast Asian language families which lack derivational morphology (Enfield 2005: 188), with some of the Mon-Khmer languages like Vietnamese shifting their typological allegiance due to areal pressure. In the database, derived terms in Sedang are virtually all restricted to the domain of artifacts; the vast majority of analyzable terms in other semantic domains are of the lexical type.

4.5.1.4. New Guinea

4.5.1.4.1. *Meyah.* In Meyah, of the East Bird's Head family, the inflection of nouns is very limited. Indeed, nouns "remain uninflected or have very little inflectional possibilities" (Gravelle 1998: 558). More precisely, animates may carry a plural suffix, and inalienable nouns obligatorily bear a prefix indicating person and number of the possessor (Gravelle 1998: 562; see Gravelle 2004: 114-130 for an overview of different types of nouns in Meyah). Verbs agree with either subject or object in person and number, and inflect for aspect,

mood and instrument (Gravelle 1998: 558). Both facts may be illustrated with the example in (36.).

(36.) *di-ra meiteb di-(e)r-agob ofa*
1-use machete 1S-INST-strike 3S
'I use a machete striking her/him'[26] (Gravelle 1998: 560)

Meyah lacks derivation as a major word-formation technique both in the verbal and nominal domain (Gravelle 1998: 564; for some minor instances of derivation see Gravelle 2004: 71). In the absence of derivation on a larger scale, verbs may be juxtaposed to specify or vary the meaning of the lexical items involved. For instance, the verb *eja* 'go' occurs in juxtaposition with adjective roots creating inchoative semantics (Gravelle 1998: 565), and similarly, in the absence of morphological causatives, *otonggum* 'make/do' and *agob* 'kill' are used to convey indirect and direct causation respectively (Gravelle 1998: 567).

Compounding operates by the juxtaposition of two lexical roots which may be of different parts of speech (Gravelle 1998: 569; see Gravelle 1998: 569-570 and Gravelle 2004: 130 for criteria for their identification). Abstract nouns in particular are formed by juxtaposition of *mar* 'thing' with other roots of different parts of speech (see also Gravelle 2004: 133):

(37.) a. *mar ɸ-oga ɸ-en-ah-ma*
thing 3S-word 3S-DUR-sits-DEM
'There is a word thing' = 'There is a dispute'

b. *ofa ɸ-odou ɸ-ebriyi gij mar moisu*
3S 3S-liver 3S-splits in thing shaman
'His liver splits in shaman things' = 'He knows about magic'
(adapted from Gravelle 1998: 570)

Example (37b.) also illustrates the frequent use of body-part terms, in particular *odou* 'liver,' for the expression of psychological states.

The formation of noun-noun compounds involving two inanimate nouns or one inanimate noun and a noun denoting a part of the body is also "highly productive" according to Gravelle (1998: 570). It is this type which makes up, in the absence of any derivational processes, the entire set of morphologically complex lexical items in all semantic domains in the database. However, N-V compounds are also attested (Gravelle 2004: 132). While complex terms consisting of two roots are most frequent, there are some items in which as many as three are combined (Gravelle 1998: 564). As noted in general by Gravelle (1998: 130), "Meyah productively uses compounding to express nominal meanings that can be glossed with one word in another language, such as English." Examples include:

[26] Additional gloss: INST 'instrument'

(38.) a. *mocgój efésa* 'fog/cloud flash' = 'lightning'
b. *méngk ofód* 'breast flood' = 'milk'
c. *mówa eitéij* 'sun eye' = 'clock'
d. *ojóna otkonú* 'married.woman stomach' = 'womb'
e. *mówa ot déis* 'sun stand middle' = 'noon'

Compounds of either type function syntactically like simplex nouns in that they may occupy argument positions, as in (39.).

(39.) *Ri-of terapan nou kabar ofu ke-uma*
3PL-cover tarp for ship fly NOM-that[27]
'They covered that plane with a tarp' (Gravelle 2004: 132)

While Meyah does feature systems of numeral and noun classification (Gravelle 2004: 137-142), these systems do not appear to be employed in a derivational fashion frequently.

4.5.1.4.2. *Toaripi.* A language with a very similar lexical profile in New Guinea is Toaripi, belonging to the small Eleman language family. Here, too, morphologically complex terms consisting of two lexical roots dominate, as seen in (40.).

(40.) a. *fave uta* 'stone/rock hole' = 'cave'
b. *ori tolo* 'bird leaf' = 'long wing- or tailfeather'
c. *a-e* 'fire-faeces' = 'spark'

Occasionally, however, one encounters complex lexical items in which the two constituents are linked by a possessive marker, such as *ofae ve mehe* 'eye POSS hair' = 'eyelash.' Also occasionally, constituents appear in reduplicated form in the morphologically complex lexical item, such as in *maea ma ma* 'sweat, perspiration' (*maea* 'body,' *ma* 'water'). Like in Meyah, artifact terms involving 'thing' are not encountered for the items on the wordlist (but given that they do exist in Meyah, they may also exist for other meanings in Toaripi); artifact terms in Toaripi are either other morphologically complex terms of the lexical kind, or are designated by semantic extension (such as *ma ove* 'water picture,' originally 'reflection in water' and then extended to 'mirror, looking-glass'), or loanwords from Motu or English.

4.5.2. COMPOSITIONAL TYPE OF POLYSYNTHESIS: KET

Vajda (2004a: 400) notes a descriptive gap in traditional morphological typology in that there is no standard term for languages with a rich inflectional, but relatively sparse derivational apparatus, in contrast to languages such as those of the Eskimoan and Wakashan families, which also have a large amount of derivational morphemes. These are prime examples of Mattissen's (2003, 2004) affixal type of polysynthesis, and are here representa-

[27] Additional Gloss: NOM 'nominalizer.'

tives of the derived type with respect to their lexicon, a fact which corresponds well with the wealth of available derivational mechanisms. Vajda suggests "conglomerating language" to describe the other language type he has in mind. His example is the Yeniseian language Ket, which is used as an example of the compositional type of polysynthesis in Mattissen (2004), meaning roughly that it allows the concatenation of more than one lexical root into the noun complex. Notably, the difference noted by both authors with regard to the positioning of Ket on the spectrum of polysynthesis when compared with languages of the Eskimoan and Wakashan families corresponds to a difference in the lexical profile for Ket in that the overwhelming majority of complex nominals in the database are of the lexical rather than of the derived type.

According to Vajda (2004a: 413), Ket favors conversion as a transcategorical, word-class changing operation. However, a nominalizing suffix *-s* does exist: examples are *nánbèt-s* 'baker' (*nánbèt*, 'to bake bread') and *úgdè-s* 'length' (*úgdè* 'long,' see also Vajda 2004b: 15 and Werner 1998: 39-41 for further discussion and examples). Vajda (2004b: 15) also mentions some archaic non-productive noun derivation devices as well as quasi-derivational use of terms for 'male' and 'female.' The nominalizing suffix *-s* is responsible for the few terms in the database classified as being of the derived type in the database, of which an exhaustive list can be found in (41.).

(41.) a. *də-doq-s* '3N-fly-NMLZ' = 'airplane, helicopter'[28]
b. *binaʁols* /bin-o-qol-s/ 'self-PST-heal-NMLZ' = 'scar'

In addition, there is redundant marking of plural on some nouns (Vajda 2004b: 20; according to Vajda 2004b: 80, "multiple plural suffixes in many nouns came to serve as stem-building elements," that is, are employed in a fashion remotely resembling derivation, see Werner 1998: 41-43 for further discussion). This phenomenon occurs for instance in *ulteɣin* /ul-te-in/ 'water-lake-PL' = 'swamp, quagmire;' a further case seems to be *cas-aŋ* 'hour/watch-PL' = 'clock, watch, timepiece.'[29]

However, compounding is a much more frequent word-formation strategy in Ket, and indeed Werner (1997: 46) states that compounding must be an old mechanism in the language since most polysyllabic and even some monosyllabic words can be traced back historically to compounds. Werner (1998: 49-50) distinguishes between proper and improper compounds ("eigentliche und uneigentliche Komposita"). For the former type, bare roots are concatenated, while for the latter type (which is particularly common if the relationship between the roots is possessive or synecdochic according to Vajda 2004b: 15), the genitive marker *-d* occurs as a linking element. Werner (1998: 50) notes that this type of construction is not typical for Yeniseian languages. Examples for compounds without such a linking element are seen in (42.).

[28] Ket also has the borrowed terms *samolop* and *eroplan* (from Russian *samolët* and *aeroplan* respectively).
[29] As Bernard Comrie (p.c.) points out, it may be relevant that in Russian *časy* 'watch, clock' is grammatically the plural of *čas* 'hour.'

(42.) a. *mamul* /maˀm-ūl/ ‘breast-water’ = ‘milk’
b. *espul* /ēs-hɯl/ ‘sky-mound’ = ‘cloud’
c. *aqqot* /ekŋ-qoˀt/ ‘thunder-path’ = ‘rainbow’
d. *eŋgaj* /eˀŋ-àj/ ‘eggs-sack’ = ‘nest’

Werner (1998: 53) notes that elements such as *-ūl* ‘water’ (as in 40a.) and *-àj* ‘sack’ in example (40d.) (as well as a number of others) may be considered to be halfway along a grammaticalization path to derivational elements, and calls them word-forming semi-affixes (“wortbildendes Halbaffix”). The reasons for such a position are that they occur in a series of compounds while at the same time diverging phonologically from their free-standing counterparts (a point also noted by Vajda 2004b: 14). However, Werner (1998: 58) states that since they can be readily identified with lexical sources, they should still be regarded as elements of a compound rather than derivational suffixes. Examples of compounds with the linking element *-d* include:

(43.) a. *bulaŋdoks* /būl-aŋ-d-ōks/ ‘leg-PL-POSS-wood’ = ‘ladder’
b. *destqār* /dēs-d-qār/ ‘eye-POSS-hair’ = ‘eyelash’
c. *sestbaˀŋ* /ses-d-baˀŋ/ ‘river-POSS-land’ = ‘coast’
d. *oləndquk* /olən-d qūk/ ‘nose-POSS hole’ = ‘nostril’

Vajda (2004b: 15) also discusses a number of examples of compounds which are unlike the examples above complex words on phonological grounds, but rather phonological phrases.

Thus, Ket, which is typologically quite different from neighboring Eurasian languages, also stands out when compared to the typically mixed lexical-derived profile of its neighbors (see § 4.6.6.), as well as by the presence of some highly complex nominals that go well beyond the combination of two roots, such as *uldʌptaŋbaŋ* /ul-də-b-taŋ-baŋ/ ‘water-3N-3N-turn-place’ = ‘whirlpool.’ Furthermore, the Ket lexicon, at least judging by the equivalents to the meanings on the wordlist used here, is characterized by an elevated degree of analyzable terms in the broader Eurasian context (in spite of Vajda’s 2004b: 14 comment that “Ket basic vocabulary includes numerous non-derived stems, many of them monosyllabic”). To conclude the discussion of Ket, the following textual example (from Vajda 2004b: 97, boldface added) illustrates a compound of the type with the genitival linking element in the context of discourse; in addition this example gives an idea of the polysynthetic nature of the Ket verb, with the morphological analysis provided in square brackets and position classes indicated by superscript numbers:

(44.) *bū* *ttíŋgòmdaq* *[du^{8}-tiŋ^{7}-k^{5}-o^{4}-b^{3}-in^{2}-daq^{0}]*
3M.SJ he.stopped.those.up [3M.SJ^{8}-plug^{7}-ADES^{5}-D^{4}-3N.O^{3}-PT^{2}-shove^{0}]

ít-dès ***d-ólàn-d*** ***qúk-sèn*** *haj* *ít-qò*
3F.POS-eyes **3F.POS-nose-N.GEN** **hole-PL** and 3F.POS-mouth
'He stopped up her eyes, **nostrils**, and mouth with sand.'[30]

Nivkh has a similar lexical profile, although there are differences in the structure of the noun complex (Mattissen 2004: 51, see the extensive discussion of Nivkh in Mattissen 2003). Also similar is the lexical profile of Cheyenne, although here a certain amount of deverbal derivation exists (see also § 4.6.2).

4.6. LANGUAGES WITH MIXED PROFILES

4.6.1. OVERVIEW

As stated in the introduction to this chapter, languages relying exclusively on complex lexical items of either the lexical or of the derived type only represent the extreme end-points of a typological continuum, and many languages (in fact, most languages) in the sample make use of both types in the creation of morphologically complex expressions and thus fall somewhere in between the extremes. This section aims to highlight some different types of languages with such mixed profiles, but it cannot do full justice to the actual diversity found. Another complication is that many, but not all, languages of this type have a very low proportion of analyzable lexical items, and thus their classification is due to a rather restricted number of examples. For instance, in Imbabura Quechua, there is an absolute number of two lexical items of the derived type in the database, which amounts to a relative number of 8.9 per cent of derived-type lexemes among the analyzable lexemes of that language. Note that in absolute numbers this is precisely the same figure as found in Ket, but due to the higher total number of analyzable terms (compounds, in this case) in Ket, it falls into the category for which the lexical type is dominant, and not in the mixed-type category. The fact that the languages in this group are so typologically heterogeneous is not least due to the comparatively low total number of analyzable terms. To illustrate this, a discussion of languages on different ends of the typological continuum (but not at the extremes) on the syntheticity-analyticity scale, Efik and Hawaiian on the one hand and Sora on the other, follows.

[30] Glosses: ADES 'adessive case suffix; in finite verbs, a derivational affix denoting motion towards,' D 'durative marker (appears in many stative and activity verbs),' F 'feminine class (a subset of animate class),' POS 'possessive clitic (proclitic on noun phrases; derived from genitive-case suffixes,' M 'masculine class,' N 'neuter (=inanimate class); either singular or plural,' O 'verb-internal direct-object agreement affix, or direct-object pronoun,' PL 'plural,' PT 'past tense,' SJ 'verb-internal subject agreement prefix, or subject pronoun.'

4.6.2. MORE OR LESS ISOLATING PROFILE: EFIK AND HAWAIIAN

A case of a mixed-type language with a relatively low degree of inflectional synthesis is Efik, a Benue-Congo language of Africa. Efik has complex expressions of the lexical type, as seen in (45.).

(45.) a. *akam'ba obüt'* 'great land/earth' = 'mountain'
b. *idib'i uküt* 'belly/womb leg' = 'calf'
c. *ök'pö ñkañ* 'bone side' = 'rib'

However, Efik seems to prefer to combine only nouns in this way. When complex items involve verbal roots, a nominalizer in the form of a prefixed vowel or syllabic nasal exhibiting a wide range of allomorphy is typically employed (Welmers 1968: 157-158, Essien 1990: 119-121), as in (46a. - 46d.). There is also the prefix *eri-,* which forms participial nouns or gerunds from verbs (46e.)

(46.) a. *e-fak'* 'NMLZ-fix.between' = 'a narrow street, a lane'
b. *e-kep'-kep* 'NMLZ-red-to.flash' = 'flashing, lightning, glistering'
c. *u-tök' ikïm* 'NMLZ-to.void urine' = 'bladder, chamber pot'
d. *ukur'isü* /u-kut-i'sü/ 'NMLZ-see-face' = 'looking glass, mirror'
e. *eritem'* 'ABSTR-cut.down' = 'clearing'

As can be seen from examples (46c.) and (46d.), these deverbal terms sometimes are accompanied by a nominal element which may be conceived of as the object of the nominalized verb (Welmers 1968: 157). The deverbal formations in general are subject to lexicalization to a large degree; it is frequently the case that the initial vowel of a noun is analyzable as a nominalizer, but the following stem does not occur outside of the construction (Welmers 1968: 157, Essien 1990: 116 on the closely related language Ibibio).

Languages with a similar profile include other African languages such as Hausa, as well as Austronesian languages such as Tetun and Hawaiian. In addition to partial and full reduplication, Hawaiian employs a variety of derivational morphemes to derive meanings from morphologically unanalyzable terms (47.). However, as seen in example (47b.), derivational suffixation may also be applied to a reduplicated form. Examples include:

(47.) a. *ʻoi-ʻoi* 'RED-sharp' = 'thorn'[31]
b. *ʻoiʻoi-na* 'pointed/protrude-NMLZ' = 'headland'
c. *kā-maʻa* 'CAUS-bind' = 'shoe'
d. *mō-lehu* 'QUAL/STAT-ashes/ash-coloured' = 'dusk'

It is perhaps noteworthy that in languages with a mixed profile and a relatively low degree of synthesis, in particular in the domain of nominals, one of the derivational processes employed is or appears to be frequently full or partial reduplication. Examples from a variety of different sample languages are:

[31] Both *ʻoiʻoi* and *ʻoi* also have other related meanings.

(48.) a. Hawaiian *lihi-lihi* 'eyelash, eyelid' (*lihi* 'edge')
b. Sahu *so-soloro* 'bird' (*soloro* 'to fly')
c. Bororo *baru-baru* 'horizon; type of spirit' (*baru* 'sky')
d. Rama *kung-kung* 'lungs' (*kung* 'louse, whit, air root')
e. Bislama *kol-kol* 'fog, mist' (*kol* 'cold')[32]

However, except for Sahu and Hawaiian, reduplication is not the dominant derivational strategy in any of the languages in this survey, and therefore there is at most a weak correlation between mixed-type languages with reduplication as a derivational device.

4.6.3. COMPOSITIONAL TYPE OF POLYSYNTHESIS: SORA

The Munda language Sora has certain typological similarities with Ket in that it belongs to the compositional type of polysynthetic languages. Details differ, however. Sora verb morphology is extensive, and Sora nouns inflect for number, person, and possession, and their function in the clause is indicated via case and case-like adpositions (Anderson and Harrison 2008: 306). In terms of word-formation, nominal roots in their combining form play a major role. Together with a large number of derivational processes including affixation and reduplication, concatenation of such nominal roots (up to three or sometimes even four roots according to Anderson and Harrison 2008: 307) serves to create the nominal lexicon of Sora, and more than one derivational element may occur in the same form (Anderson and Harrison 2008: 307, 327; see also Anderson and Harrison 2008: 330 for a "monster noun form" featuring four derivational infixes). Combining forms may also be incorporated into the verb complex (most often, the incorporate corresponds to the patient of a transitive verb, Anderson and Harrison 2008: 355), and even here, more than one combining form may be concatenated, as example (49.), from Anderson and Harrison (2008: 359) who quote Ramamurti (1931: 143), shows:

(49.) *jo-me-bo:b-dem-te-n-ai*
smear-oil-head-RFLXV-NPST-ITR-1
'I will anoint myself with oil'[33]

Examples of complex nouns of a variety of types in Sora are seen in (50.). These include concatenation of combining forms (50a. - 50d.), sometimes including additional material such as a reflexive morpheme in (50b.).

[32] Both *kolkol* and *kol* also have other meanings.
[33] Additional glosses: ITR 'intransitive,' NPST 'non-past,' RFLXV 'reflexive.'

(50.) a. *əsu:ŋtidən* /ə-'su:ŋ-'tid-ən/ 'POSS-hut.for.temporary.use-bird-N.SFX' = 'nest'[34]
b. *duŋdəm'da:(ba:)n* /duŋ-dəm-dˀa:-(ba:)-n
'get.out.of-RFLXV-water-(place)-N.SFX' = 'a natural spring of water'
c. *'kuru:tamən* /'kuru·-'tam-ən/ 'body.hair-mouth-N.SFX' = 'beard'
d. *dai'jʊŋən* /dai-jʊŋ-ən/ 'climb-sun-N.SFX' = 'sunrise'

4.6.4. DERIVED TERMS IN THE DOMAIN OF ARTIFACTS AND STRUCTURAL HOMO- AND HETEROGENEITY IN DERIVATIONAL DEVICES

In spite of the large structural heterogeneity of languages of the mixed type noted above, it is still possible to discern certain patterns in their behavior. In some languages of this type, derived terms for the meanings under investigation are very common or even found exclusively in the domain of artifacts, but are rare or unattested in the other domains. This is true for instance in Chickasaw, in which two thirds of derived terms cluster together in the domain of artifacts. The two most common basic schemes are locative nominalization with the prefix *aa-* and the nominalizing suffix *-'*, and usage of *isht* 'with' together with the same nominalizing suffix *-'* for instrument nominalization. As the example in (51b.) makes clear, the basic types may be amended by additional bound morphological material:

(51.) a. *aa-nosi-'* 'LOC-sleep-NMLZ' = 'bed'
b. *aa-ili-pisa-'* 'LOC-REFLX-see-NMLZ' = 'mirror'
c. *ishtalakchi'* /isht talakchi-'/ 'with be.tied.up-NMLZ' = 'rope'

Santiago Mexquititlan Otomí is a mirror image of this in that derived terms are not found in the domain of artifacts at all (the language has largely borrowed its vocabulary in this domain from Spanish), but are dominant in the domain of nature-related lexical items:

(52.) a. *mbo-zaa* 'place.where.around-tree' = 'forest'
b. *ma-hets'i* 'LOC-high' = 'sky'
c. *h-ñä* 'IMPERSONAL.VOICE-speak' = 'steam, voice, language'
d. *munts'i* /m-punts'i/ 'NMLZ-to.relapse' = 'whirlpool'

Similarly, in Wichí, analyzable terms in the domain of nature-related terms are almost exclusively of the lexical type (53a., 53b.), with the exception of the term for 'rain,' which consists of a subject marker prefixed to the verb root *-wu-* 'to make' and followed by an epenthetic morpheme and a locative suffix (53c.).

(53.) a. *itoj muwk* /itoj mukw\n/ 'fire dust' = 'ashes'
b. *tewuk lhip* 'river part' = 'coast'
c. *i-wu-m-cho*' '3SBJ-make-EP-LOC.under' = 'rain, to rain'

[34] Gloss in the source is 'a nest of binds' [sic].

In contrast, Wichí terms for artifacts and body-parts are often of a different structure; some of them involve prefixation of the possessive marker *to- ~ tot-*. The terms of the derived type in this language occur frequently in combination with this possessive marker, as seen in the examples in (54a., 54b.), (although there are also complex terms of the lexical type with this prefix as well as artifact and body-part terms which do not feature it, like for instance 54c.).

(54.) a. *tot-telhu-hi-s* 'POSS.INDET-eyes-LOC.in-PL' = 'glasses'
b. *to-nhes-pe'* 'POSS.INDET-nose-LOC.on' = 'nostril'
c. *y'amekw-hi* 'excrement-LOC.in' = 'toilet'

San Mateo del Mar Huave is a Mesoamerican language that behaves similarly.

Furthermore, another independent variable is that, in some of the sampled languages, derived terms are mostly construed by use of one and the same derivational marker. This is the case in Chickasaw, for which the majority of such terms involve the nominalizer *-'* (often in conjunction with *aa-* 'locative' or *isht* 'with'), while in other languages, such as Santiago Mexquititlan Otomí, terms classified as being of the derived type are of a variety of structural types.

Table 2 cross-classifies languages with mixed profiles according to these two variables. The criterion for assignment is whether 50% of the derived terms are in the domain of artifacts or not, and whether 50% of them are formed by the same derivational device or not. This is not always possible, for instance when the number of derived terms is equally distributed over the artifact domain and other domains or when the number of derived terms is precisely one. Languages in which one of these situations obtains are not shown in the table. For languages in which one derivational device dominates, the corresponding morphemes and glosses are given in parentheses.

	different devices	mostly same device (given in brackets)
mostly artifacts	4 Nuuchahnulth, Carrier, Basque, Bororo	4 Hausa (*ma-...-i* 'INSTR-...-INSTR'), Chickasaw (*-'* '-NMLZ'), Haida (*-ʔu* '-INSTR'), Guaraní (*-ha* '-AGT'),
mostly or exclusively non-artifacts	16 Great Andamanese, Pipil, Arabela, Kiliwa, Kiowa, Hawaiian, Welsh, Kanuri, Pawnee, Jarawara, Cavineña, Santiago Mexquititlan Otomí, Itzaj, Imbabura Quechua, Embera, Bislama	4 Sahu (reduplication), Sora (*-ən* '-N.SFX'), Kolyma Yukaghir (*-i:* '-INSTR'), San Mateo del Mar Huave (*-aran* '-INAL.POSS')

table 2: languages with a mixture of complex terms of the derived and lexical type and their structural diversity and spread over semantic domains

As is clear from the table, there is a strong skewing for derived terms to not be in the domain of artifacts in languages with many different devices to form complex words of the derived type. Specifically, there are only four languages in which these are found domi-

nantly in the artifact domain; in contrast, in the vast majority of such languages, the terms are more frequent in other semantic domains. Among them are Pipil, Santiago Mexquititan Otomí, and Itzaj, which have borrowed terms for artifacts from contact languages, in most cases Spanish, rather than coining neologisms (§ 5.4.2.7.1. deals with borrowing vs. coinage of neologisms, the influence of the contact language as well as structural factors in more detail), but also Kiliwa and Kiowa, which did the precise opposite. At any rate, there is a skewing with respect to languages with high structural diversity in derived terms when cross-classified with the semantic domains these are found in that is not observable in languages where one structural type dominates.

Apart from this classification, there are also commonalities shared between some languages with a mixed profile. Whether these stand in a direct relationship to their assignment to the mixed type is a question that must remain open for the time being; nonetheless, these commonalities by themselves are striking enough and of interest for lexical typology that they are worth discussing. The first commonality is areally and perhaps, on a more subtle level, even structurally motivated, and pertains to complex nominals with a clause-like structure in many sampled languages of North America (§ 4.6.5.). The other is areal only and stems from the observation that many languages of Eurasia rely on a mixture of derived and lexical terms in their morphologically complex terms (§ 4.6.6.).

4.6.5. Verb-centered Nominals in North America (and beyond)

4.6.5.1. Overview

Mithun (1999: 287) notes for North American languages in general in the context of the discussion on ceremonial speech that "[t]he structure[s] of many North American languages ... lend themselves well to such periphrastic expression, due to their general polysynthesis. Single words are often composed of many meaningful parts, and their literal meanings are in many cases still perceptible to speakers. Indeed, words in all domains are frequently coined from complex descriptions ..." While complex nominals in North America have a wide range of structural types, as seen above, there is one particular type which is quite common in languages spoken in North America, although the phenomenon in question is neither found in all languages of the continent nor does it need to be the primary type for the formation of complex nominals. The type in question are clausal nominals that are formally either similar to headless or internally headed relative clauses or are fully inflected verb forms which may sometimes, given the appropriate context, receive a compositional interpretation as such, but which have specialized institutionalized semantics to denote a particular object or entity. Languages in which either of these are found typically fall into the mixed category due to the mixture of verb-based derivation involving a single root and the clausal structure involved in other terms involving more than two roots. Such languages have no clear lexical profile. In spite of this, the similarity of these devices is evident enough to justify the postulation of a separate type of word-building, due also to it appearing as either areally or typologically restricted.

4.6.5.2. Clausal Nominals with Relative Clause-like structure

4.6.5.2.1. *Kiliwa.* Kiliwa, a Yuman language of Baja California, is a good example for illustrating nominals with relative clause-like structure, as well as for showing that this strate-

gy typically coexists with other types of word-formation, which together account for the morphologically complex expressions in the lexicon. Kiliwa nominals, for instance, may be compounds that can either have nouns or adjectives as constituents:

(55.) a. *nmi ʔ=tay* 'cat=large' = 'cougar; puma'
b. *haʔ=ñmiʔ* 'mouth=fur' = 'moustache; beard' (Mixco 2000: 25)

More importantly, however, there are mechanisms that allow for the formation of deverbal nouns, of which Kiliwa has a wide variety. As discussed by Mixco (1965: 100), there are traces of archaic and non-productive types of nominalization that highlight the essentially verb-oriented character of the language, such as a vowel shortening rule evidenced by some striking resemblances of some nouns to verb roots, e.g. *n-míʔ* 'cat' (cf. *(ʔ)-mîˑ* 'to cry') and *ʔ-wáʔ* 'house' (cf. *wâˑ* 'to dwell). Further evidence for a verbal origin of a large number of Kiliwa nouns is provided by nouns which seem to be analyzable into a verbal root and a verb prefix, thus not showing any sort of overt marking for nominalization, such as *č-pàt* 'doorway' (*č-* 'perpendicular axis prefix found on verbs,' *pàˑ* 'to leave'), *m-phú* 'hole' (*m-* 'stative,' *-phú* 'to perforate,' Mixco 1965: 93-94). Furthermore, there is evidence for lexicalization of erstwhile nominal compounds with a phonological contraction in the first element. Mixco (1965: 106) derives the first element *w-* of *w-xâˑ* 'cave' and *w-ñáʔ* 'road' diachronically from **wîˑy* 'mountain, stone(?).' In fact, Mixco (1965: 92) characterizes the entire nominal lexicon of Kiliwa by saying that "[n]ouns either give the impression of being eroded nominal compounds or atrophied verbs, with fossilized affixation and traces of vestigial verbal processes such as reduplication and vocalic ablaut."

Synchronically analyzable morphemes to derive nouns from verbs include:

(i) *t-* 'reifier,' as in *t-híp* 'grass, fodder' (*-híp* 'to sprout') (Mixco 1965: 95-96; glossed OBJ in Mixco 1985)
(ii) *-uʔ* 'locative/nominalizer.' Terms derived with this suffix frequently have a locative meaning, as for instance in *h-pâˑ-uʔ* 'sleeping area' *(-pâˑ* 'to lie down'), but it is also employed as a general nominalizer (Mixco 1965: 96). Derivations with this suffix may in turn enter into larger structures such as nominal compounds, such as *mñis w-uʔ* 'foetus sit-LOC' 'womb' (a term taken from the database)
(iii) *-tay*. This postclitic is identical to the verb *-tay* 'to be large' and has a nominalizing function as in *hphíʔ táy* 'musical instrument' (*phíʔ* 'to make sound'), with the resulting noun denoting someone or something that performs the action of the verb it is derived from frequently (Mixco 1965: 97).

However, the most productive morpheme in the language in general according to Mixco (1965: 95) is the prefix *k*^{w}-, which derives nouns from adjectives with the meaning 'one characterized by x' or 'one that does x,' where x is the meaning of the verb root, as in the examples in (56.).

.

(56.) a. *phi? k^{w}-msúx* 'nostril' = 'nose which is perforated' (Mixco 1965: 102)
b. *k^{w}-phú·y* 'smoke, grey' (*-phû·y* 'be smoky') (Mixco 1965: 95)

Nominalization with *k^{w}-* is also the most interesting process from a typological point of view since *k^{w}-* also functions as a relativizer. In the following more complex examples (from Mixco 2000: 25), the overlap in function between nominalization and relativization becomes particularly clear. All examples involve a further nominal root which is modified by a nominalized verb by means of *k^{w}-*, or, since the process of nominalization and relativization appears to be indistinct in these cases, a relative clause introduced by *k^{w}-*.[35] Mixco (1965: 101) calls such structures "relativized compounds."

(57.) a. *xaq-m=k^{w}-sk^{w}aayu* 'beef-OBJ=WH-guard' = 'cowherd'
b. *myal=k^{w}-sn?aawu* 'tortilla=WH-small:PL' = 'cracker; cookie'
c. *?wa?=k^{w}-s?hin* 'house=WH-run' = 'automobile'
d. *xwa=?iy=kw-kuus* 'enemy=hair=WH-long:PL' = 'Cocopa'

But there is more to be said about deverbal derivation in Kiliwa. Notably, two or more nominalizers may be combined. Mixco (1965: 97) calls this characteristic "complex nominalization." Examples are given in (58.), with the derivational morphemes highlighted by boldface (see also Mixco 1965: 97-98 for examples 58a. and 58b.)

(58.) a. ***t-kw-****ipaa-y* 'OBJ-WH+be.alive-ATT' = 'animal'
b. ***t-****xpapu-****u?*** 'OBJ- sew-OBL' = 'needle'
c. *wa?=t-****kw-****lkwii-y-****tay*** 'house-OBJ-WH-carry-ATT-FREQ' = 'train'

Furthermore, there is a type of relativized compound in which there is no overt mark of nominalization according to Mixco (1965: 101). Examples provided by Mixco (1965: 102) only include terms for animals, such as *?múw há? ñmî? k^{w}yúw* 'sheep (which has) (on the) mouth hair which stands' = 'goat,' so it is unclear just how productive this type of nominalization is in Kiliwa.

Kiliwa compounds readily combine with bound nominal morphology, such as the demonstrative and illative suffixes *-mi* and *-l* in example (59a.). In this example, the compound seems to be in apposition to the actual clause syntactically. However, Kiliwa compounds also readily occupy argument positions, as seen in (59b.), where the compound is marked as subject by the suffix *-t*:

(59). a. *mi=?-m-mi-l* *č?ii-k,* ***mîy=ha?****-mi-l*
this=DN-COM-this-IL stick-DIR, **leg=face**-this-IL
'An arrow stuck right here in his **calf** [Achilles spot].'
(Mixco 2000: 58, boldface added)

[35] According to Comrie and Thompson (2007: 379), indistinctiveness of nominalization and relativization is found in a number of languages in Western and Southwestern North America (see further references there).

b. *yu-m=yuu-t, h-paa* ***yuw=xaʔ****-t* *xwaq-uʔ-l* *xiʔwap-m-t*
be-DS=be-SS, 3-recline **eye=water**-SBJ two-REL-IL spill-DIR-SS
'And so, as he lay weeping, his **tears** spilled on either side
(of the watershed)'[36] (Mixco 2000: 52, boldface added)

It would be very interesting to know how other types of complex nominals, in particular those involving *kʷ-* behave syntactically, but unfortunately, no such example occurs in the short text samples provided in Mixco (2000).

4.6.5.2.2. *Cheyenne.* In Cheyenne, the marker *tsé-*, glossed in the source as 'that which is,' plays an important role, in particular in complex lexemes in the domain of nature-related terms. It may be prefixed to a verb root to form the name for a topological feature, as in (60a.). Its apparent relativizing function becomes clear by the fact that verbs prefixed with *tsé-* may be accompanied by a simplex noun that it modifies, with the derived verb either following (60b.) or preceding the noun (60c.).

(60.) a. *tsévé'evótoo'e* /tsé-vé'evótoo'e/
'that.which.is-be.a.concave.hole' = 'cave'
b. *tsésééha mahpe* /tsé-séeha mâhpeve/
'that.which.is-spread.out water' = 'lake'
c. *hotohke tséana'ôhtse* /hotohke tsé-ana'ôhtse/
'star that.which.is-fall' = 'meteoroid'

However, in the domain of artifacts, unlike in some other languages, the construction apparently plays a minor role. Here complex formations with *hestôtse* 'thing' dominate:

(61.) a. *táxemésêhestôtse* /táxe-mésehe-hestôtse/
'upon-eat-thing' = 'table'
b. *ameohestôtse* /ameohe-hestôtse/
'go.by.quickly-thing' = 'car'

In addition, Cheyenne also features a large number of compounds, such as *háhnoma-pano'êhasëö'o* /háhnomâ-hépano'êhasëö'o/ 'bee-syrup' = 'honey' (compounds in fact outnumber the deverbal type of complex lexemes), leading to the coexistence of many different structural types of morphologically complex expressions in the lexicon.

4.6.5.2.3. *Carrier.* In Carrier, a candidate for a morpheme which both has a relativizing function and is used to coin complex nominals is *-î,* glossed here as 'REL' in the following instances.

[36] Glosses: COM 'commitative [sic!] (with),' DS 'different subject,' IL 'illative (in; inside),' REL 'non-subject relativizer,' SS 'same subject.'

(62.) a. *hananelyih-î* 'grow.again-REL' = 'seed'[37]
b. *dzeł-î-krez* 'mountains-REL-between' = 'valley'

However, there appear to be other structural types of verb-based nominals. Morice (1932: 84-85) points to terms such as *uḳwetsežthi* 'bed,' literally translated as 'on it one lies down' and *uṭaz teratsiyaih* 'ladder,' literally 'by the help of it one goes up.' These are, in the words of Morice (1932: 84), "nothing but a verb preceded by a pronominal preposition without a substantive," and it is these examples which provide the transition to another structural type of complex nouns found in some languages of the world, in particular those of North America: derived nouns without any overt indication of the derivational nominalizing process.

4.6.5.3. Nominalization without Nominalizing Morphology

One phenomenon encountered frequently in languages of North America concerns expressions with nominal meanings which are morphosyntactically similar or identical to finite verb forms. This type of construction is also the structural phenomenon that underlies personal names such as 'dances with wolves,' and it will thus be referred to as the 'dances with wolves'-type of nominal mnemonically. It is "a commonplace for Americanists" according to Beck (2005: 3), and indeed, such constructions are found in a number of sampled languages.

4.6.5.3.1. *Kashaya.* In Kashaya, derived nominals with the same inflection as found on verbs may be formed by a morpheme called the absolutive (this is an entirely different type of marker than the eponymous absolutive marking S and P in languages with ergative alignment, and should not be confused with it). This suffix, which has a wide range of phonologically conditioned surface allomorphy, is the citation form for verbs. In narratives, the absolutive form of the verb is the most frequent form of the main verb (Oswalt 1961: 266). In other words, a verb inflected with the absolutive is clearly finite. However, verbs inflected with the absolutive also have a nominal function, and "[t]he absolutive behaves syntactically just like a common noun" (Oswalt 1961: 266). For instance, as seen in example (63.) from the database, it may enter into larger constructions with other nouns.

(63.) *qʰaˑmoś šuṭʰuhṭadu* /qʰaˑmoś šu-hṭʰuṭ-ci^{O}d-w/
'star by.pulling-pieces.come.off.bigger.object-DUR-ABS' = 'shooting star'

This is the only example in the database in which the absolutive morpheme is found in terms from the domain of nature and topology. It is more frequent in terms for artifacts, particularly when combined with a following element *=li* (here tentatively analyzed as a postclitic, but possibly an affix or a postposition), indicating instrument or location.

[37] The "literal" translation provided by the lexicographer for *hananelyih-î* is "that which uses to [sic!] grow again." Presumably "uses to" is meant to highlight the habituality; however, the present discussion is only concerned with the fact that the literal translation employs a relative clause.

(64.) a. *caduli* /cad-w=li/ 'see-ABS=INSTR' = 'mirror'
b. *caq^{h}awli* /ca-q^{h}a-w=li/ 'WITH.MASSIVE.OBJECT-cut.off-ABS=INSTR' = 'knife'

Interestingly, Oswalt (1961: 267) reports that the plural of *ciwalaw* 'shirt,' which is derived from a complex verb meaning 'to crawl down' by means of the absolutive suffix (Oswalt's "literal" translation is 'crawling down'), was remembered by a consultant as having been *ciyalaw*, which incorporates additional verbal inflection for pluractionality, thus meaning something like 'crawling down repeatedly' when interpreted verbally. However, at the time of Oswalt's field work this form had been displaced by inflection with the regular nominal plural *ciwalá·ya*. Oswalt's (1961: 267) conclusion is that "[t]he Absolutive form may be used so predominantly as a noun that there is a tendency to lose sight of its verbal origin."

4.6.5.3.2. *Oneida.* There are four categories of nouns in Oneida (Abbott 2000: 47-48): (i) monomorphemic nouns with no internal structure, which are small in number, (ii) nouns consisting of a nominal root which require a prefix and a suffix that do not add any discernible meaning to the root, but instead simply serve to identify the forms in question as nouns, (iii) deverbal nouns derived from verb roots by suffixation of a nominalizing suffix (and in turn followed by a noun suffix), and (iv) "syntactic nouns," which are "formed by constructing a verb, typically with an indefinite pronominal prefix and serial suffix, and then simply using that verb syntactically as a noun" (Abbott 2000: 48). According to Michelson (1990: 76), "the majority of functional nominals are verbal derivatives or forms which are structurally indistinguishable from verb forms."

(65.) a. *kahnekóni?* / ka-hnek-No-ʔ/
'NEUT.AGENT-liquid/liqour-be.in.water/cook.in.water- STAT'
= 'wells, puddles'
b. *tetwaʔsʌ́thaʔ* /te-w-aʔsʌt-haʔ/
'DUALIC-NEUT.AGENT-drop:CISLOCATIVE-HAB'
= 'waterfalls, Niagara falls'
c. *tekaˑtʌ́heʔ* /te-ka-tʌ-heʔ/
'DUALIC-NEUT.AGENT-fly-HAB' = 'airplane'

In principle, the fourth way of forming nouns results in ambiguity as to their interpretation, that is, the forms may be interpreted either as verbs expressing actions or as nouns denoting entities (Michelson 1990: 77). However, Michelson (1990: 77-80) also gives a number of examples for different structural types in which the nominal reading is associated with a particular set of inflections, and when these are varied, the resulting interpretation is verbal and no longer nominal. More generally, she notes throughout that there is variation in how freely speakers provide "literal" meanings for the individual forms in question, which is further evidence that the forms are lexicalized with the nominal reading to different degrees.

In the domain of artifacts, it is particularly common to find the habitual aspect suffix combined with an instrumental suffix. As seen in examples (66c.) and (66d.), this combination of morphemes is also found in a number of body-part terms:

(66.) a. *yelathʌstákhwaʔ* /ye-lathʌst-hkw-waʔ/
'FEM.INDEF.SG.AGENT-get.something.up-INSTR-HAB' = 'ladder'
b. *yehyatúkhwaʔ* /ye-hyatu-hkw-aʔ/
'FEM.INDEF.SG.AGENT-write-INSTR-HAB' = 'pencil, pen'
c. *yewelaláhkwaʔ* /ye-wel-l-hkw-aʔ/
'FEM.INDEF.SG.AGENT-air/wind-be.in.or.on-INSTR-HAB' = 'lungs'
d. *-khwalákhwaʔ* /-khw-l-hkw-haʔ/
'-food-be.in-INSTR-HAB' = 'my stomach'

According to Abbott (2000: 53), one characteristic of Oneida terms referring to land forms is that they are either incorporated into a verb indicating position or stance, or they carry a locative suffix.[38] An example of the former construction from the database seems to be (67.):

(67.) *kélhiteʔ* /ke-lh-Nit-eʔ/
'ANOMALOUS.PREFIX-woods-be.in[39]-STAT' = 'forest, trees'

4.6.5.3.3. *Pawnee.* According to Parks (1974: 90), compounding is the most widely employed word-formation technique in Pawnee (see Parks 1974: 123-139 for a typology), although nominalization, which is at the focus of attention in this discussion, is also important.[40]

Nominal inflectional morphology in Pawnee is very sparse: there are locative, instrumental and diminutive suffixes; inflection for number is restricted to certain contexts, and only kinship terms are marked for (inalienable) possession. When not followed by any of the suffixes, and when not part of a larger construction such as a compound, nouns appear with the so-called "absolutive suffix" *-u* (Parks 1974: 97-98; this suffix appears as *-uʔ* in the lexical source). Other nominal stems cannot occur independently except when occurring with the diminutive *-kis*, of which Parks (1974: 94) says that it often seems to function as a plain nominal suffix without diminutive semantics. An example from the database would be *rikucki*, underlyingly /rikuc-kis/ 'bird-DIM' = 'bird' (compare Parks 1974: 103).

Nominalization in Pawnee occurs in a variety of subtypes, but all of them employ the subordinate form of the aspectual suffixes (subordination and aspect are intertwined in Pawnee, see Parks 1974: 188 for details). An example provided by Parks for a nominalization from an active verb (denoting actions and processes) involving the subordinate

[38] Parallels are also found in Tuscarora. For instance, the root *-(ę)ʔty-* 'bay' only occurs incorporated in the verb *-yę(T)-* 'to lay.'
[39] This root is featured in the term only "possibly" according to the consulted source.
[40] These statements are mirrored in the typological position of Pawnee: the majority of complex nouns in Pawnee are classified as being of the lexical type, and a minority as being of the derived type.

imperfective marker *-hus* is *kickawiriwu*, Parks's (1974: 117) literal translation of which is "paddles the water" (note the relevance of third person agreement being realized by zero). This is commonly used in terms for items of acculturation:

(68.) a. *ukaʔaatawiiriku ~ ukaatawiiriku* /ukaʔaata-wi-iirik-hus/
'be.a.shadow-LOC-see-IPFV = 'mirror, glass, window'
b. *rakcaakarikuku* /rak-caakarikuk-hus/ 'tree/wood-unlock-IPFV' = 'key'
c. *raktariihkaruukus* /rak-rariihkaraʔuk-hus/ 'tree/wood-write-IPFV' = 'pen'

Note that there is some variation as to whether the surface forms retain the final *s* of the aspectual suffix (Parks 1974: 117). The same is true of another type of Pawnee nominalization, the "gerundial" type (Parks 1974: 120; Parks and Pratt 2008 gloss the relevant prefix, *ra-*, as 'iterative'). Of this type, only one example is found in the database, *ra-raar-awarii* 'ITER-ITER-fly.around' = 'airplane;' the literal translation provided for this is 'one flying around.' Other gerundial nominalizations mentioned by Parks (1974: 120) also feature an aspectual suffix (see also Parks 1974: 121-122 for further structural variants of the gerundial type).

A further noteworthy feature of Pawnee is that a considerable number of terms for topological features and natural kinds lack a nominal equivalent entirely. Instead, the meanings are encoded formally as verbs. Examples include, but are not limited to:

(69.) a. *huu-kiihaar* 'enclosure-on.a.surface'
= 'be a valley, be a bottoms, as of a stream course'
b. *kic-takaahak* 'water-pass.down'
= 'water to drip, be a waterfall, be a rapid, rapids, as in a stream'
c. *tat-kus* 'stalk-to.be.sitting' = 'be a plant; plant to be (growing)'
d. *waa-wikiiʔac* 'hill-to.be.growing' = 'be a mountain, be a tall or high hill'

Another type that also occurs in other semantic domains consists of a verbal root with the nominal suffix *-uʔ:*

(70.) a. *kiwahaar-uʔ* 'be.a.lake-NOM'
= 'pond, lake, slough, body of water, area of standing water'
b. *awiriituʔ* /awirit-uʔ/ 'be.hot-NOM' = 'heat, steam'
c. *ratkahaar-uʔ* 'be.night-NOM' = 'night'

These are all pieces of evidence for a verb-centered nature of the Pawnee lexicon (see § 5.4.1. for further discussion with reference to Pawnee).

4.6.5.3.4. *Nuuchahnulth.* Relying on a large set of suffixes for the purpose of word-formation, this Wakashan language lacks compounding entirely according to Davidson (2002). Examples from the database are in (71.), see also Davidson (2002: 193-195) for an overview of devices and further examples.

(71.) a. *hita-ċus* 'LOC-dig' = 'bay'
b. *ƛ̓ap-ẏak* 'straddle-INSTR' = 'scissors'
c. *hapuuɬ* /hap-quw̌ɬʷ/ 'hair-on.the.face' = 'beard'
d. *ċaċałaqṅukum* /ča-ċałaq-nukʷ-im/
'RED-branch.out-in.or.at.the.hand' = 'finger'

Such formations can become more complex, as the examples in (72.) show.

(72.) a. *čaapaciiɬmaʕuk*
čapac-i:ɬ-maʕuk
canoe-making-one.skilled.in
'one who is skilled in making a canoe' (Nakayama 2001: 65)

b. *ċuuƛ̓ċuuyaẏak*
REDUP-ču-(y)a-ẏakʷ
ITER-wash-REP-instrument.for[41]
'basin' (Nakayama 2001: 66)

Nakayama (2001: 68) provides a detailed discussion of the scalar nature of conventionalization and lexicalization, which is similar to that noted for Greenlandic by Fortescue (1980), and of the varying degree to which Nuuchahnulth speakers are aware of the morphological structure of complex words, depending to some extent on the suffix they contain.

However, Nuuchahnulth also features clausal nominals with finite verb morphology. These are called "nouns derived from lexicalized verbs" by Davidson (2002: 328). Among the examples given by him are those in (73.).

(73.) a. *ƛ̓ʷiti·ƛ̓ʷitš* 'sticks on at intervals' = 'hummingbird'
(*ƛ̓ʷiT-* 'stick on' + iterative I aspect)
b. *qʷa·ya·ċikši·ɬ* 'Turns-into-Wolf' = '(man's) name' (*qʷayaċi:k* 'wolf' "with perfective (and hence verbalizing) suffix *-šiƛ̓* and iterative II aspect")

As Davidson (2002: 328) also notes, this type of nominal is a "particularly common source for personal and place names," and this fact explains their absence in the data for the present study. Notably, Nuuchahnulth is (in-)famous for an alleged lack of a distinction between noun and verb (a claim going back to Sapir 1911), and indeed Nakayama (2001: 44) emphasizes that Nuuchahnulth stems are generally able to predicate, regardless of whether they have a more "nouny" or "verby" semantics (compare the further discussion on distinguishing word classes in Nakayama 2001: 41-57). Nakayama (2001: 57) and Davidson (2002: 325) seem to converge in the position that the distinction between noun and verb

[41] Glosses : ITER 'iterative,' REP 'repetitive.'

may be made, but that it is grammaticalized only to a low degree, and of course, this fact is potentially highly relevant for the presence of nominals with finite verbal morphology.

The textual example in (74.) shows a derived noun for 'animal' embedded in natural discourse, and the derived noun in (75.) does the same for an expression meaning 'village' that is verbal in nature, as evidenced by the presence of inflectional morphology for mood (emphasis added in both cases).

(74.) *qʷiinuuƛ̓ḥiič* *wiiy̓aat*
qʷi-a·nu:ƛ̓-(q)ḥ-(y)i:-č wi:y̓a-'at
that.which-because.of-SIM-INDF.3-INF.3 never-SHIFT
why.you.would never

haẇaaḥin?at *saštup*
haẇ-a·ḥin-'at sa-štu·p
eat-deprive.of-SHIFT **crawl-species**
take.food.away.from animal
'The reason why you never take food away from **animals**.'[42]
(Nakayama 2003: 262)

(75.) *naʔaa ʔan ciiqciiqa ḥaa quuʔasm̓inḥʔi čaakupiiḥ*
naʔa· ʔan ciq-(y)a ḥa: qu:ʔas-m̓inḥ-ʔi· ča:kupi:ḥ
hear that speak-REP there human-PL-DEF men
hear that speaking there the.people men

ʔanč wim̓aaqƛ̓at ƛ̓awiičiʔat hiy̓atʔitq
ʔan-č wim̓a:qƛ̓-'at ƛ̓awa-'i·čiƛ̓-'at hiɬ-'atḥ-ʔi·tq
that-INF.3 incapable.of-SHIFT near-INC-SHIFT there-residing REL.3
that cannot.be.done get.close **where.they.lived**

qʷayaċiikštaqumɬ
qʷayaċi:k-štaqumɬ
wolf-groups
wolf.tribe
'He could hear the people saying that it was impossible to get close to the wolf **village**'[43] (Nakayama 2003: 157)

4.6.5.3.5. *Blackfoot.* In Blackfoot (Algonquoian), a similar but not quite identical situation is encountered. Here, it is possible to use intransitive verb stems with the appropriate verbal inflection as nouns by "reclassification" (Frantz 1991: 116), as in the following examples:

[42] Glosses: INDF 'indefinite mood,' INF 'inferential mood,' SHIFT 'perspective shifting,' SIM 'simultaneous ('while doing ...).'
[43] Additional glosses: INC 'inceptive aspect," REL 'relative mood,' REP 'repetitive aspect.'

(76.) a. *káta'yáípasskaawa* /káta'-á-ipasskaa-wa/
= 'NEG-DUR-dance-3S' = 'non-dancer'
b. *áókstakiwa* /á-okstaki-wa/
= 'DUR-read-3S' = 'reader' (Frantz 1991: 117, small caps added)

One difference to, for instance, Pawnee is that the third person affix is not zero, so that the clausal structure of these nominals is fully brought to light. As Frantz (1991: 116) points out, there is evidence for reclassification when these forms are pluralized: they take the plural suffix *-iksi*, which is used for animate gender nouns, but in the singular, the forms are indeed ambiguous since *-wa* occurs on both nouns and verbs.

Blackfoot, however, also features other types of nominalization. Deverbal nominals involve affixation of verbal markers indicating person and number that are "nearly identical" to those of an inflectional paradigm (the so-called conjunctive paradigm) found on verbs in subordinate clauses (Frantz 1991: 120). This is remarkably similar to the nominalizations found in Pawnee. As Frantz further points out, "[i]t is probably more realistic to speak of this phenomenon as clause nominalization, for not only do the verbs agree with subject or object, but all other elements which normally accompany verbs in clauses may be present with Conjunctive Nominals" (emphasis removed). However, it is important to emphasize that, unlike in other North American languages already discussed, the nominalized nature of such clauses is indicated by the presence of a nominal suffix following the verbal inflection from the subjunctive paradigm. The main types of nominals formed in this fashion are locational nominals, temporal nominals, instrumental nominals, and manner nominals, each indicated by different prefixes (Frantz 1991: 121-124). The most important constructions for the present study are locational nominals, formed by the prefixation of *it- ~ iit-*, and instrumental nominals, formed by the prefixation of *omoht- ~ iiht- ~ oht-*. Examples (from Frantz 1991: 121-122) include:

(77.) a. *iitáísóooyo'pi* /iit-á-iso-ooyi-o'p-yi/
'there-DUR-on-eat-21:NOM-IN.S' ='table (what one eats upon)'[44]
b. *iitáóoyo'pi* /iit-á-ooyi-o'p-yi/
'there-DUR-eat-21:NOM-IN.S' = 'where one eats/restaurant'
c. *kitsítáóoyihpoaawayi* /kit-it-á-ooyi-hp-ooawa-yi/
'2-there-DUR-eat(AI)-NOM-2P-IN.S'
= 'where you$_{2p}$ eat/your restaurant'
d. *iihtáóhpommao'pa* /iiht-á-ohpommaa-o'p-wa/
'INSTR-DUR-buy-21:NOM-3S' = 'money/what one buys with'

Comparing examples (77b.) and (77c.) indicates that they are apparently conventionalized as their meaning may be modified by the affixation of additional morphemes.

4.6.5.3.6. *Biloxi.* The "literal translations" offered in the consulted source for Biloxi suggest that clausal nominals with no nominalizing morphology are also present in the Siouan

[44] Glosses: DUR 'durative,' IN.S 'inanimate singular,' NOM 'nominal suffix.'

language Biloxi, in particular in neologisms in the domain of artifacts, as in the examples presented in (78.), which are presented together with the "literal" translation from the source.

(78.) a. *ina ´do^{n}-hi ´* 'clock' ("sees the sun")
b. *akŭtxyi´ on´ni ~ akŭtxyi´ on* 'pen, pencil' ("makes writing" or "makes books")
c. *o^{n} ´do^{n}ho^{n}ni ´* 'mirror' ("what is used for looking at or seeing")

The precise nature and extend of the phenomenon in Biloxi, however, cannot be determined on the basis of the extant sources.

4.6.5.4. Discussion: An areal or typological phenomenon?

The overall areal distribution of such formations remains at present somewhat unclear, although it is obvious that they are well entrenched in North America. However, Hagège (1993: 174) points out that "[m]ost languages can use whole sentences as if they were a single unit which functions like a noun" and cites English *forget-me-not* and French *m'as-tu-vu* 'lit. did you see me' = 'one who likes to show off' (next to the Mbum term for 'pinky' already mentioned in Hagege 1970) as examples. Given this evidence, the question one should probably ask about clause-like nominals appears to be one of quantity, that is, in which languages they form a regular mechanism of lexical enrichment and what enables them to do so rather than to ask for the necessary condition for such formations per se.

While the phenomenon is clearly well represented in North America, Beck (2005: 3) reports the following data showing a lexicalized noun with the structure of a relative clause from Upper Nexaca Totonac in Mesoamerica:

(79.) *ti:* *ki-ma:-w-í:*
HREL 1OBJ-CS-eat-CS[45]
'my wife' (lit. '3SG who feeds me')

It may be that kin terms, which are inherently relational and in this feature are unusual as nouns, are particularly amenable concepts to be expressed by such constructions. Thus it is unclear how far the phenomenon extends southwards in the Americas. Likewise, nominals with the surface structure of finite clauses are by no means restricted to North America. There are some particularly interesting cases reported from polysynthetic languages of Australia. In Bininj Gun-Wok, there is a highly polysynthetic verb complex with twelve prefix and two suffix slots (Evans 2004: 71), while the nominal morphology is restricted to noun class prefixes on some nouns and optional case suffixes for non-core grammatical roles (Evans 2004: 76). There is productive compounding, but with semantic restrictions as to which nouns are permissible elements of compounds. For instance, *bininj* 'man' or *daluk* 'woman' cannot be involved in compounding, and as a result compounds making reference to the occupation or typical activity of individuals are rendered impossible. Instead, "[c]oncepts of this type are frequently expressed as deverbal nominals, in which a fully

[45] Glosses: CS 'causative,' HREL 'human relative pronoun,' OBJ 'object.'

inflected word is simply used as a syntactic noun without any overt signaling of class-change; these may or may not have an incorporated nominal" (Evans 2004: 98). Examples can be found in (80.), also from Evans (2004: 98):

(80.) a. *ka-warlbu-n* '3-hunt-NP' = 'he hunts'/'hunter'
b. *kabarri-bolk-nahna-n* '3a-land-look.after-NP'[46]
'they look after the land' = 'land owners, custodians'

For Yir Yoront, Alpher (1991: 72) likewise reports the existence of "[n]ominalized clauses with no nominalizing morphology" that involve finite verbs. These are not very frequent, however, and are restricted to terms referring to animals and plants. Mithun (1984: 507) quotes the the example in (81.) from Berndt and Berndt (1951: 34-35) from Guwinggu, another Australian language which shows the same or at least a very similar structure:

(81.) *gandijigar'garmere gadbere ŋa:d nawu garigugbu´lere*
they keep it for us ours us these our bodies are dark
'they look after our (language) for us Aborigines.'

An obvious question that arises is: what is it that enables many languages of the Americas (and others, apparently particularly those of Australia) to coin words with clausal structures or at least resembling clausal structures to a considerably greater extent than, for instance, better known European languages? And, equally important, what allows these languages to not only coin such words, but also to have them enter the lexicon as conventionalized terms for a given referent? One factor that at least appears to play a certain role is head-marking clausal morphology for the conventionalization of the relative-clause-type of morphologically complex items exemplified by the Kiliwa terms constructed using *k*^{w}-. Although Kiliwa does have some elements of dependent-marking, including number marking (Mixco 2000: 15) and optional though common case marking indicating the grammatical function of NPs in the clause (Mixco 1965: 110), "[t]he noun phrase in Kiliwa is considerably simpler in its morphology than the verb phrase" (Mixco 2000: 24). A further relevant factor appears to be that the Kiliwa relativizer *k*^{w}- is merely an affix (on the verb, i.e., a head-marking element), which perhaps facilitates the conventionalization of complex word forms by virtue of the fact that, as an affix, it is a simple fixed unit within a lexical item, thus avoiding the structural clutter involved in other strategies. For instance, it is hard to imagine that a language with relative pronouns such as those found in languages of Europe would coin a complex description analogous to Kiliwa *(?)wa?kws?hin* 'house which runs' = 'automobile' that would then enter the lexicon as a fixed expression; rather, if at all, such a language likely would prefer other strategies of word-formation such as compounding to express the same content.

As for complex lexemes of the 'dances with wolves'-type, the discussion here must remain even more speculative, but an obvious area of grammar that is worth looking at is syntactic organization. One line of argumentation that is already found in Boas (1911)

[46] There is no index of abbreviations and glosses in Evans (2004).

claims that in some languages of North America overt noun phrases do not have argument status or that this status is at least questionable (see Milewski 1950 for an early comparative treatment of phrasal structure in languages of North America). Instead it is argued that they are in apposition to the pronominal affixes found on verbs, which constitute the "real" arguments of the verb (Jelinek 1984, cf. also Kibrik 1992: 137). This entails the verb opening up slots for arguments that are however immediately saturated by the pronominal affixes attached to them. This so-called pronominal argument hypothesis, however, is highly controversial. Connected to this question, but equally controversial, is the claim made for some languages of North America that they lack a lexically anchored distinction between nouns and verbs. For instance, this is argued for by Sasse (1993) concerning the Iroquoian language Cayuga; one piece of evidence provided is the existence of clausal nominals as in the discussion on Oneida above. Mithun (2000), in contrast, rejects this position, arguing that nouns and verbs can be distinguished by clear criteria in these languages, while still admitting that forms classified on morphological criteria as verbs may function syntactically as nouns. In her view, this fact does not entail the lack of a possible distinction between noun and verb (in both authors' discussions, the existence of nominals with verbal inflection is not the only piece of evidence adduced). For the purpose of the present study, it is neither necessary to accept or reject the pronominal argument hypothesis, nor to take a position on the question of the distinction between noun and verb in Iroquoian and other languages. With regard to the former, it is possible to at least note that noun phrases in Oneida and Pawnee, two of the languages in which the 'dances with wolves'-type of nominal is frequent, do not bear any overt signaling of their function in the clause, i.e. case (this is also true of Nuuchahnulth). This, to be sure, does not entail noun phrases not being arguments, but at least it is not incompatible with an interpretation of noun phrases as appositions rather than arguments. Still, it must be emphasized that, as seen above, there are structural differences in the formation of clausal nominals, and it would be necessary to assess the syntactic organization for each individual language in more detail before well-founded broad generalizations are possible. Looking at the relationship between referential and predicative expressions (as done by Sasse 1991) and the structure of discourse is perhaps also revealing, and will be done here using textual data from Oneida as an example.

In this study, individual lexical items are primarily considered in isolation, separated from the context they occur in. This is to some extent unavoidable, but one must not forget that their use by actual speakers is embedded into larger structures of discourse, which is in turn related to the overall lexicogrammatic system of the individual languages. Examining how the types of complex nominals are integrated into discourse is particularly instructive for languages such as Oneida, in which types of morphologically complex items exist that differ quite radically from garden-variety word-formation devices such as compounding and derivation. Consider the following excerpts from an Oneida story:

(82.) a. *nʌ ne? nʌ?n*
P P P

s-a-yakwa-an-itskw-A-hl-A-?tsl-ot-u-nyu-hE?
ITER-AOR-1EXPS-SRF-**thigh-JOIN-set-JOIN-NOM**-stand-DIST-DIST-SER
again we set up the **chairs**

tsi?	*yo-atek-ha?*	*wa?-t-yakwa-hwanhak-E?*
P	NO-burn-SER	AOR-DU-1EXPS-tie.up-PUNC
where	it (the fire) is burning	we tied around

'Then we would put the **chairs** all around the fire'[47]

(Michelson 1981: 13, 38-9, boldface added)

From the discourse-pragmatic point of view, reference is being made to 'chairs' on which something is predicated in this passage. The term for 'chair' in Oneida is morphologically complex, but conventionalized as a fixed expression, occurring in two alternants. When free-standing, its form is *anitskwahlákhwa?*. This is a morphologically complex artifact term of the type discussed in §4.6.5.3.2., featuring the instrument suffix *-hkw* followed by the habitual suffix *-ha?*. However, in Oneida, this type of "verby" nominal has the ability to be incorporated into a verb, in which case the sequence mentioned above is replaced by the nominalizing morpheme *-?tsl* (Michelson 1990: 76; this seems to be in contrast to other Iroquoian languages such as Cayuga, in which only monomorphemic nominal roots have the ability to act as incorporate, cf. Sasse 1993). It is this second alternant of the word for 'chair' that appears in the above passage incorporated into the verb root *-ot-* 'to stand,' which is in turn inflected multiple times with morphemes conveying mostly aspectual information, effectively creating a highly complex verb form consisting of a total of thirteen morphemes. What can also be seen in the above example is that the discourse in Oneida is heavily centered around the morphosyntactically defined verb (and presumably, Oneida would score quite low on referential density indices such as that developed by Bickel 2003). Although the discourse functions reference and predication can clearly be identified on the level of the sentence, morphosyntactically speaking, the stretch of discourse consists of a series of predicates realized by verb forms which are interspersed with unanalyzable discourse particles such as *nʌ, ne?, nʌ?n,* and *tsi?.* In other words, discourse is heavily verb-centered. In the following excerpt from the same story which further illustrates the nature of Oneida discourse structures, a clausal nominal with verbal inflection meaning 'quilt' occurs in an unincorporated form.[48]

[47] Glosses: 1EXPS '1exclusive plural subjective,' AOR 'aorist mode,' DIST 'distributive,' DU 'dualic,' JOIN 'stem-joiner,' NO 'neuter objective,' NOM 'nominalizer,' P 'syntactic particle,' PUNC 'punctual aspect,' SER 'serial aspect,' SRF 'semi-reflexive.'

[48] To be sure, these examples were selected to illustrate the types of constructions in Oneida that are of interest for the present study. As noted in the discussion, there are also other types of nouns in Oneida, such as those consisting of a root plus noun prefix and suffix, and of course they also figure in actual discourse. For instance, in the following excerpt, the noun root *-aht-* 'shoe' has a prefix indicating a female possessor and the noun suffix *-a?.*

b. *s-w-at-yel-ʌ* *kʌs* *yo-thole-ʔ* *thikʌ nʌs kwi*
ITER-NS-SRF-do-PERF P NO-be cold-PERF P P P
sometimes it is cold weather

te-ka-ihn-A-kkhani-ʔ *t-a-yukhi-yu-ʔ*
DU-NS-cloth-**JOIN-piece.together-PERF** CIS-AOR-F/ldp-give-PUNC
cloth pieced together she gave it to us

neʔ *waʔ-t-yakwa-at-haʔuwEʔek-Eʔ*
P AOR-DU-1EXPS-SRF-wrap.up-PUNC
this we wrapped it around us

'Sometimes it would be kind of cold and she would give us each a **quilt** to wrap around in'[49]

(Michelson 1981: 13, 39, boldface added)

One can again observe the same structures: a series of verbal forms (including that for the referential expression 'quilt') interspersed with a number of discourse particles. Similar remarks are also made by Mithun (1984: 505) for Mohawk, another Iroquoian language: "Normal discourse consists predominantly of morphological verbs, since verbs can function not only as clauses and predicates, but also as nominals, with no modification in form. Such nominals are often verbal descriptions of their referents, like *ra'swà:tha'* 'he extinguishes'/ 'fireman'" (see also further discussion in Mithun 1984 as to the role of lexicalization and to what extent such forms can be regarded as subordinate). The following example, which seems quite similar to the structures found in Oneida, gives a taste of this type of discourse organization in Mohawk:

(83.) *Ó:nen ki' ne rahtahkón:nis tahoná:khwe'.*
now just the he shoes makes he got mad
'At this point the shoemaker became angry'

(Mithun 1984: 506, quoting Phillips p.c.)

What is the relevance of this for the phenomenon at hand? The details of syntax and their interaction with the lexicon in Oneida (and Mohawk, and also other languages in which similar structures are found) clearly require detailed study by experts to elucidate the intricate relationship between grammar and lexicon. Nonetheless, it seems safe to at least

(i) *yah teʔ-w-eʔni-Ø* *n-aʔ-te-yo-stalathe-ʔ* *yako-aht-aʔ*
NEG NEG-NS-evident-PERF PART-AOR-DU-NO-shiny-PUNC FPOSS-shoe-NSF
'Her shoes were really shiny' (Michelson 1981: 25, 49)

Glosses: PART 'partitive,' FPOSS 'feminine-indefinite,' NSF 'noun suffix'

[49]Additional Glosses: CIS 'cislocative,' DP 'non-singular,' F/ldp 'feminine non-singular' (the sequence /l is unclear), NS 'neuter subjective.'

claim that the verb-centered nature of discourse allows for the smooth integration of clausal nominals, either in the form of relative clauses or, as is the case here, in the form of finite verb forms, into the overall organization of discourse. This statement, to be sure, is not an explanation of the phenomenon per se, but at least it helps to understand how the structures in question fit into the overall lexicogrammatical system of the language.

4.6.6. MIXED LANGUAGES OF EURASIA

4.6.6.1. Overview

Languages of Eurasia, including the Caucasus, have, as will be seen later, a comparatively high degree of unanalyzable, monomorphemic lexemes (attributed by Sasse 2001: 503 to the atypical millennia-long history of borrowing). With respect to the nature of their complex lexemes, it is also noteworthy that they are consistently inconsistent, that is, they mostly rely on a mixture of the derived and lexical strategy.

4.6.6.2. Greek

Modern Greek is typical in this respect. In this language, derived-type analyzable terms are found next to lexical-type analyzable terms, and it has been chosen to illustrate the typical mixture of the two types in Eurasia because many loanwords found throughout Eurasia are ultimately of Greek origin (as shown by example 84a. and b.), and while they are unanalyzable in the borrowing languages, they are analyzable in Greek. Examples of the different types (see Joseph and Philippaki-Warburton 1987: 216-229 for an overview of word-formation in Greek) are:

(84.) a. *oríz-ōn* 'bound/delimit-NOM.SG.NEUT' = 'horizon'
b. *mîlon toú Adám* 'apple-NOM.SG.NEUT ART.GEN.SG.MASC Adam' = 'Adam's apple'
c. *lik-ó-fōs* 'wolf-STEM.FORMATIVE-light' = 'dawn'

As examples (84b. and 84c.) also shows, complex lexemes may be of the compounding or phrasal type in Greek.

4.6.6.3. Laz

Another Eurasian language with this profile is Laz, a Kartvelian language. Analyzable terms are relatively few in number; if they occur, the lexical type is somewhat more frequent, but the derived type is also well attested:

(85.) a. *kinçi-toma* 'bird-hair' = 'feather'
b. *ɜ'ari-gza* 'water-way' = 'river'
c. *o-ťoç-aṣe* 'DERIV-throw/shoot-DERIV' = 'weapon'
d. *o-bere* 'LOC-child' = 'womb'

While Laz has been in contact with Greek, it should be emphasized that the mixture of the derived and lexical type is frequent in Eurasia generally, in particular Western Eurasia. A mixed profile is also found in Basque, Bezhta, Sora, Welsh, and Kolyma Yukaghir. Likewise, Khalkha lexical items may be of both types as well (though here derived terms are a little

more numerous than in the aforementioned languages), and in Kildin Saami, analyzable lexemes in the database are of the lexical type, though generally Saami languages also feature a fair amount of derivatives.

4.7. CHAPTER SUMMARY AND A FIRST TYPOLOGICAL CORRELATE

In the above, a coarse first step was taken to the typologization of the dominant types of analyzable lexemes found in the world's languages. The only factor involved so far is the relative degree of analyzable terms of the lexical and derived type, and two extremes of the continuum on which individual languages may be placed were defined. For illustration, data from languages with different typological profiles and their interaction with the grammatical organization were presented. While this embryonic typology cannot (and is not intended to) capture particulars of all languages in the sample and does not assign every language to a particular type, it clearly does reveal an association of preferred word-formation techniques with certain linguistic types. It was pointed out that the areal distribution of those languages is skewed, with languages in which the derived type is dominant clustering in certain areas of the world, most notably the American Northwest (affixal type of polysynthesis) and the Northwest Amazon region in South America (derivational use of noun classifiers). However, as a first step, already at this stage of the investigation a striking grammatical correlate of this typology can be identified. In the discussion above, it was noted that several authors pointed to a predilection of so-called isolating languages for the lexical type, and an isolating language was defined rather loosely with reference to Greenberg (1960) as one with a low ratio of bound to free morphemes. However, it can of course also be defined with reference to individual grammatical properties. For instance, in a prototypically isolating language, arguments will not be marked for the type of relation they bear to the verb (case), and the verb in turn will not cross-reference properties of the arguments by means of affixes, while Vajda (2004: 421endnote3) notes that "[l]anguages with polypersonal verbs (i.e. verbs that internally cross-reference one or more syntactic terms or semantic roles in addition to the subject or agent) are normally regarded as polysynthetic on this basis alone."[50] Typically, this cross-referencing involves indication of person, and thus data from Siewierska (2005) on the occurrence of verbal person marking provide one opportunity to empirically measure and test the manifest hypothesis of whether there is a correlation between inflectional possibilities and a preference for either the derived or lexical type (note that this is only one test case, and the degree of synthesis cannot be exclusively defined by this measure, to be sure; examining many others categories in a similar fashion would be possible and indeed worthwhile). Siewierska distinguishes five types of languages: (i) those which do not have person marking, (ii) those that mark only the A argument, (iii) those that mark only the P argument, (iv) those that mark either A or P, and finally (v) those that mark both A and P. Since types (iii) and (iv) are quite rare cross-linguistically, the typology was simplified by removing these categories; additional data from the consulted grammatical sources was gathered for

[50] Whether or not this view is justified is another question; this statement is merely meant to illustrate the relevance of person marking as one factor for the isolating-(poly)synthetic distinction.

the languages in the statistics sample to fill gaps in the data because Siewierska's sample and the present sample do not always overlap (data are in Appendix C).

To ensure that the areal clusters of the types identified in § 4.3. do not interfere with the testing of this hypothesis, advanced statistical techniques are called for. These are introduced here, and will be used throughout the following chapters in various places. Generally, the data were analyzed by using Linear Mixed Effect Models (see Winter 2011 for a practical introduction) using R code by Bates and Maechler (2009) and Baayen (2009). This method of data analysis is becoming more and more important in psycholinguistics and phonetics research, when it is important to exclude unwanted factors, such as variation between test subjects, which have the potential to bias the outcome. The variable or variables of interest which one expects to have a systematic effect (in other words, which one believes to be a predictor) on some other variable (here, the percentage of derived terms) is typically included in the model as a so-called fixed effect, while a variable potentially having a random or unpredictable effect which one wants to control for is included as a so-called random effect (hence the term Mixed Model).

In this case, the variable of interest as a predictor is elaborateness of verbal person marking, and hence it is included in the model as a fixed effect, while area, in ensuring that it does not distort the obtained picture, is included as a random effect (see Cysouw 2010 for a similar approach in linguistic typology using Generalized Linear Mixed Models). As a rule, the Dryer-6 breakdown will be used exclusively to control for area in statistical modeling since it is the best-established of the different macro-areal regions of the world presently in use. The other two breakdowns as established by Nichols will be used in addition when areal effects on the typological variables under survey are of interest themselves in elucidating patterns both on a larger and a smaller scale that otherwise remain obscure under the Dryer-6 breakdown.

Normality and homogeneity of the model were then checked by visual inspections of histograms of residuals and plots of residuals against fitted values (and in addition, for further similar analyses, a Shapiro-Wilk test for normality and/or a correlation test was carried out between fitted and residual values if visual inspection leaves doubts as to the normality of the data). In this particular case, the response variable, i.e. the degree of derived terms, was subject to square transformation to ensure that the data are roughly normally distributed (this is always indicated when done in further tests to follow). All these are standard procedures with this type of statistical analysis to ensure that the observed effect really is genuine. Modeling was begun by including both a random intercepts component (which would correspond to differences in the value of the response in different areas, here, the areal asymmetries in the percentage of derived terms detected in § 4.3) as well as a random slopes component (which would correspond to differences in the impact of the predictor variable on the response in different areas, here, for instance, a strong effect of verbal person marking on the percentage of derived terms in one as opposed to a weak or even non-existent one in another area). Subsequently, a likelihood ratio test was carried out to ascertain whether the random intercepts component is needed and was removed if the likelihood ratio test did not reveal a significant difference between the models including and excluding it, which was the case for this particular model (unless otherwise indicated in the discussion of subsequent chapters, it was the case that

the random slopes component was not needed). The presented *p*-values are estimated by Markov Chain Monte Carlo simulations with 10,000 replicates (in case the resulting value for further models is between .03 and .07, that is, closely around the significance threshold, a further simulation with 100,000 replicates will be carried out to obtain a more reliable estimate).

A word of caution is necessary at this point: one of the most prominent traditional goals of linguistic typology is the identification of language universals and the determination of what is and what is not a possible human language. In doing so, many researchers not only seek to determine properties that living languages have, but also want to make inferences about any language ever produced by human beings in the past and any language that ever will be produced by human beings in the future. As pointed out by Maslova (2000), this poses a problem for any statistical analysis of the data, since the languages spoken today, from which such features can be extrapolated, represent only a small subset of the larger set of all languages ever produced, and their distribution is dependent to some degree on historical contingencies. Since no measures to remedy this fact (see Bickel 2008 for proposals) are applied here, the claims about correlations in the present sample are "universal" only in the sense that they extrapolate on the present-day (or near past) linguistic diversity in the world, and no claims are made that they constitute design features of all possible human languages (following Cysouw 2010: 258fn5 in this).

Finally, the result of the analysis is as follows: independently of area, there is a very significant correlation (p = .0022) between verbal person marking and the relative percentage of derived terms in the languages under investigation. As the plot in figure 6 shows, the derived type is least well represented in languages with no person marking, more frequent in languages in which the A is cross-referenced on the verb, and by far most frequent in poly-personal languages which cross-reference person information of both A and P on the verb.

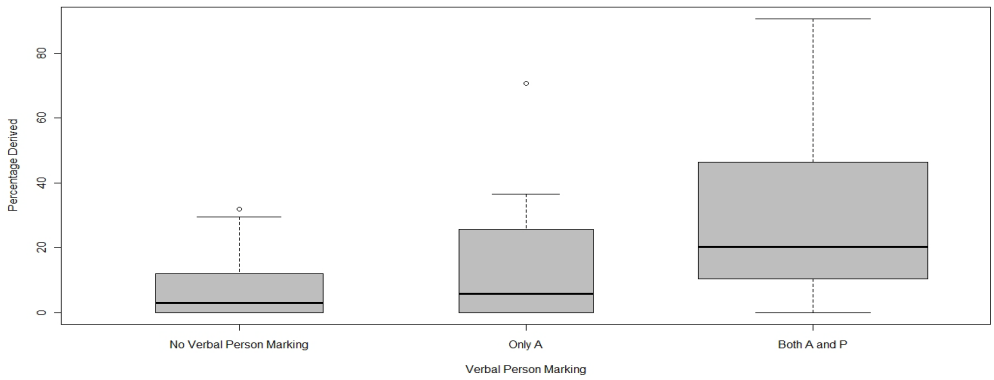


fig. 6: correlation between the preponderance of derived terms and verbal person marking

Consequently, even at this stage, a typological correlate to the classification can be added: THE MORE INFORMATION REGARDING PERSON IS CROSS-REFERENCED ON THE VERB, THE HIGHER THE AMOUNT OF TERMS OF THE DERIVED TYPE FOUND IN THE LEXICON. This fact can be interpreted as evidence in favor of the statistical correlation between an "isolating" typology and the dominance of compounding as the main word-formation strategy (as well as evidence in favor of a correlation between an elaborate verbal morphology with the derived type). [51]

Nonetheless, the above typology does not explain one of the major factors of interest for this comparative study of analyzable terms in the vocabulary, namely the differential degree to which the lexicon actually features such terms: from only a very few in some languages to quite impressively large numbers in others. There is no statistical effect at all of a language's preference for analyzable terms of the derived or lexical type on the actual number of these that are found in the lexicon ($p = .8610$ by the same Mixed Model design with percentage of derived terms as the response square-transformed). Thus, languages of both types may feature either a rather small or a comparatively high number of analyzable lexical items, regardless of their type. This means that a new dimension must be added to this typology, and other factors that influence this variable must be sought. In addition to considering other quantitative evaluations of the observed data, this is the question that will be the main focus of the following chapter.

[51] But note that there is no evidence for a correlation between the presence of case-marking for core arguments with the placement of languages on the derived-lexical continuum.

Chapter 5

Results I: Quantitative Evaluation

5.1. INTRODUCTION

This chapter is concerned with quantitative aspects of lexical motivation. It seeks to explain the behavior of the languages of the world with respect to their characteristic lexical profiles by asking questions such as: which languages have many morphologically complex terms and why? Are there languages that prefer metaphor-driven conceptualizations over contiguity-driven ones and why?

Since these are essentially quantitative questions, quantitative methods to analyze the data are needed, and therefore this chapter will make heavy use of statistics to come up with valid infererences and cross-linguistic generalizations. A concomitant and probably unavoidable effect is that for each language mostly abstract numbers rather than concrete lexical items, which ultimately are what can be and what is observed, will be analyzed statistically. In other words, there is a danger of tinkering statistically with numbers whose connection to the properties of actual languages is sometimes rather hard to see. This possible impression will be countered by making ample use of case studies that tie the data and the observed correlations to actual synchronic or diachronic observations about the languages in question to make the findings more palpable to the reader, and more generally to avoid the danger of an unduly abstract feel of quantitative analysis. Still, this chapter is characterized by quantitative methodology and probabilistic statements. The following chapter six, which is concerned with individual meanings and the cross-linguistic properties of the terms expressing them, will have a less quantitative and more of an anthropological orientation, and it will use ample data from individual languages.

5.2. DEGREE OF ANALYZABILITY: BASIC ANALYSES

In this section, the discussion of the different degrees of analyzable lexical items found in the languages of the world is entered. To give a first impression of the variability found here, the map in figure 1 shows the relative percentage of morphologically complex expressions of all languages in the core sample.

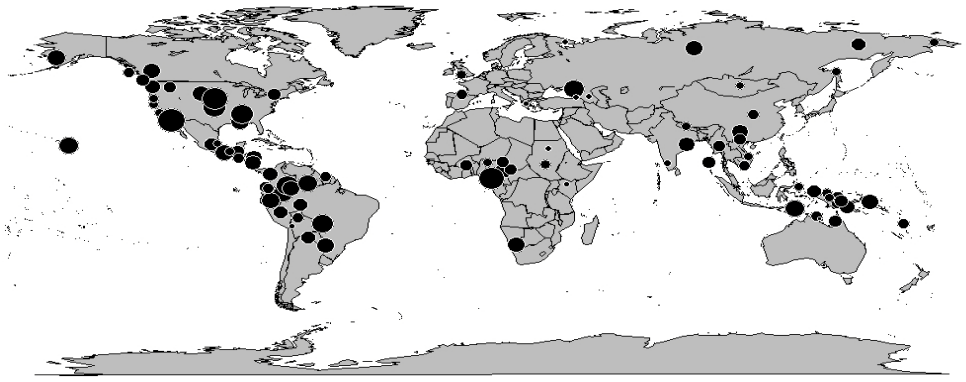


fig. 1: percentage of morphologically complex terms, core sample

The eye-catching areal clusterings will be discussed in § 5.3. But first, a basic comparison of the obtained values with that from another source, namely the World Loanword Database, follows as a kind of reliability check.

5.2.1. COMPARISON WITH THE DATA FROM THE WORLD LOANWORD DATABASE (HASPELMATH AND TADMOR 2009C)

The World Loanword Database (Haspelmath and Tadmor 2009c) contains vocabularies with about 1,000-2,000 entries for 41 languages of the world; the choice of meanings is based on, but not identical with Buck (1949). As the editors themselves note, there is a bias in the data towards European languages and thus the choice of languages is not necessarily representative of cross-linguistic diversity. The goal of the project is a systematic investigation of borrowability in different semantic domains and the varying degree of loanwords in different languages. Along with information on the status of each individual lexical item with respect to borrowing, contributors were asked to systematically code whether the lexical items are morphologically complex and, if they are, to provide a morphological analysis. This offers a convenient possibility for comparing the results of both investigations for each of the meanings that figure in both projects. This is measured by Haspelmath and Tadmor's "simplicity score," which is computed somewhat differently and thus requires some transformation to make the values comparable. The simplicity score, as the name suggests, measures morphological simplicity as opposed to morphological complexity, which is why it was converted into a measure of complexity for present purposes by subtracting the simplicity score from one. Further, the simplicity score of an individual lexical item is defined as being 1 for unanalyzable lexemes, .75 for semianalyzable lexemes and .5 for analyzable ones. To account for the difference in the scales and to convert the results into percentages the resuling value was multiplied with 200; in summary, the formula for converting simplicity scores is 200 × (1-simplicity score). Note, however, that the difference with respect to semianalyzable lexemes remains, since

they are assigned an intermediate value by Haspelmath and Tadmor while they are not taken into account in the present study at all.

129 of the 160 meanings presently investigated are also found in the World Loanword database, and the data show that the values of the different studies are often in close agreement to one another. Figure 2 is a scatterplot of the values obtained from both studies (data are in appendix D) that shows this correlation visually and is thus more accessible. Meanings with translational equivalents for only two or less languages in Haspelmath and Tadmor (2009c) were ignored in this plot; the values for 'lightning' and 'bolt of lightning' were averaged to a value of 43. As immediately becomes clear, there is a strong correlation: on average, the higher the value of morphological complexity for a given meaning in the present study, the higher the modified simplicity score value from Haspelmath and Tadmor (2009c). Unfortuntately, statistical testing is not permitted in this case, because the data overlap in some cases as some data from the World Loanword Database are included in the present sample. However, the close agreement between the two samples is very unlikely to be entirely caused by this data overlap, since the large majority of data in the sample for the present study do not come from the World Loanword Database.

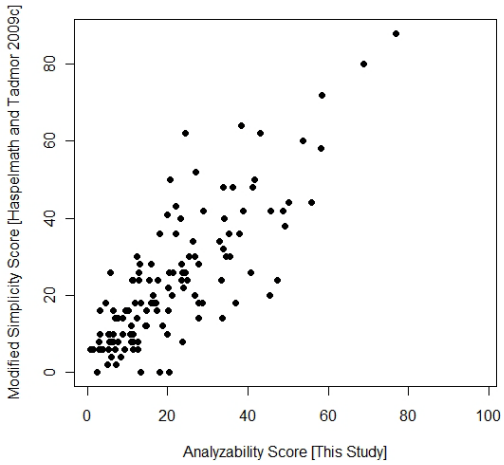


fig. 2: Correlation between modified simplicity scores from Haspelmath and Tadmor (2009c) and the measure of morphological complexity for each meaning that both studies investigate.

A further implication of this correlation pertains to the reliability of the data of this study: the vocabularies in the World Loanword Database are exclusively provided by experts on the respective languages, and thus it can be expected that all instances of morphological complexity were identified by the authors for the 41 languages in the database. As noted in Chapter 3 on methodology, mistakes in the recognition of morphological complexity in the present study cannot be ruled out and are indeed likely to occur to some degree, due to the necessary evil that the data are obtained mostly from secondary sources such as

dictionaries. However, the close agreement to the values derived from data provided by experts is a strong hint towards the assumption that overall, the data of the present study are by and large reliable.

Comparison with the World Loanword Database is also interesting in another respect: the wordlist of the World Loanword Database is considerably larger than the 160-item list of the present study. While the latter list, in spite of its relatively small size, offers a principled comparison of the sampled languages, it is nevertheless interesting to ask to what degree the values obtained from evaluating the data gathered for this list are representative for the situation with respect to the content-word inventory of the languages as a whole, i.e. including the verbal domain. Is the degree of analyzability here similar to that observed in the nominal domain? To tackle this question, Bradley Taylor has kindly computed the simplity score as defined in Haspelmath and Tadmor (2009c) for each of the languages in the World Loanword database by dividing the sum of simplicity scores for each word in a language by the sum of all words for that language. This value was transformed with the same formula as used above, and the resulting data are in Appendix B.

Notable is that the values from the World Loanword Database are significantly higher overall. This is likely a result of the fact that here many less "basic" meanings are taken into consideration which is a rough confirmation of the intuition that lexemes in more specialized vocabulary areas are more often morphologically complex than words for relatively basic concepts. However, while the values are consistently higher, the degree to which they are so vary significantly, from the rather modest difference of 5.22% in Takia up to more than 36% in Gurindji.[1] Figure 3 plots the results.

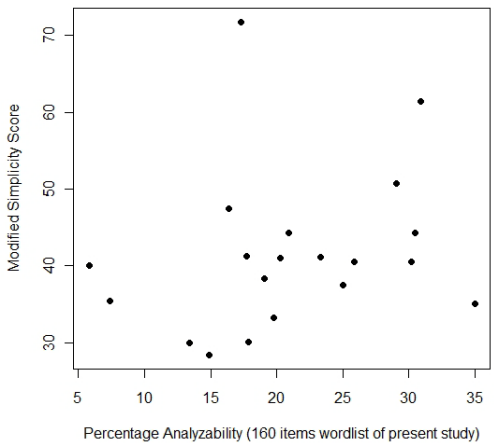


fig. 3: Correlation between modified simplicity scores from Haspelmath and Tadmor (2009c) and the measure of morphological complexity for each language that both studies investigate.

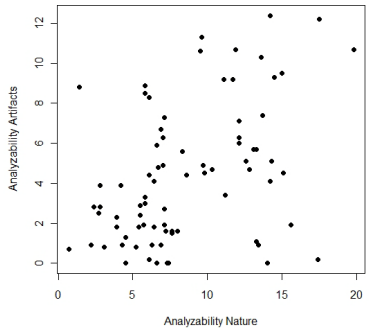
[1] Note that the even more extreme difference seen in the data for Mandarin are primarily due to the high amount of semantically redundant compounds, which are disregarded here, but which are counted in Haspelmath and Tadmor (2009c).

Again, since in this case the data for the 160-item list are a subset of the much larger overall vocabulary, statistical testing is not permitted. Figure 3 reveals a slight upward trend in the overall simplicity score as morphological complexity in the 160-items list increases, but there appears to be no strong dependency between the variables. In a way, this result is unsurprising, given cross-linguistic differences in the complexity of nouns and verbs reported e.g. for Kalam by Pawley (1993) and the typology of verbally and nominally oriented languages outlined by Talmy (2000: 59endnote11).

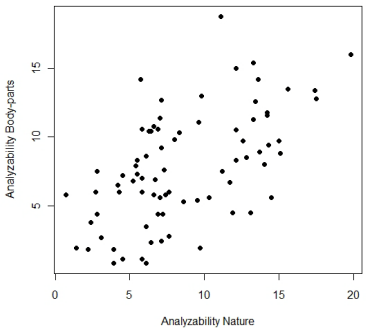
5.2.2. ARE THERE SIGNIFICANT DIFFERENCES IN ANALYZABILITY BETWEEN LANGUAGES ACROSS SEMANTIC DOMAINS?

It is conceivable that a language may rely on morphologically complex expressions in one semantic domain, while having an essentially unanalyzable lexicon in another. Data for percentages of analyzability for each language in the sample assessed over all semantic domains is in Appendix B, where information as to how the global value is distributed over the individual semantic domains is also provided (slight deviations from the global value are due to rounding).

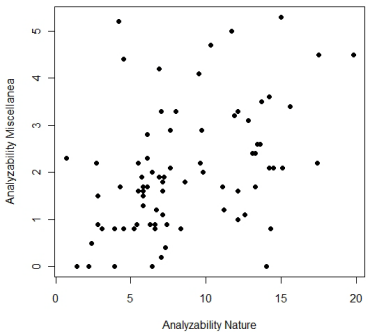
The question how morphological complexity is distributed across domains can be statistically assessed by performing correlation tests for each semantic domain with the others on the basis of the statistics sample. The diagrams in figure 4 plot the correlation between analyzablity in the four semantic domains; a correlation measure (Spearman's ρ) and an approximate *p*-value (due to ties) is provided in addition. The reported *p*-values are adjusted using the Bonferroni correction as implemented in R because of multiple testing.



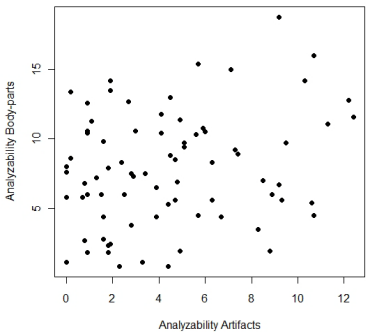
Nature vs. Artifacts: ρ ≈ .40, $p < .002$



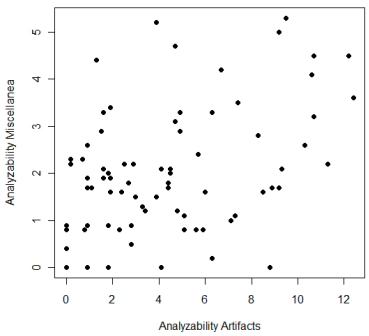
Nature vs. Bodyparts: ρ ≈ .54, $p < .0001$



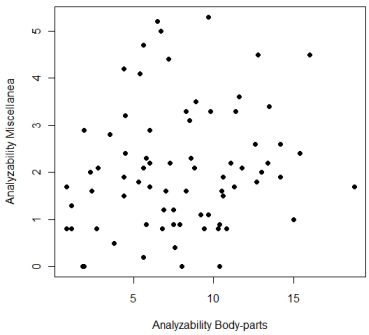
Nature vs. Miscellanea: $\rho \approx .44$, $p < .0004$



Artifacts vs. Bodyparts: $\rho \approx .20$, $p \approx .47$



Artifacts vs. Miscellanea: $\rho \approx .36$, $p < .007$



Bodyparts vs. Miscellanea: $\rho \approx .23$, $p \approx .28$

fig. 4: correlations in analyzability between meanings in different semantic domains

While there clearly is variation between semantic domains (this pertains in particular to the domain of artifacts and bodyparts), there are no dramatic cross-linguistic differences. The generalization that emerges is that THE ANALYZABLE TERMS ARE ON AVERAGE DISTRIBUTED FAIRLY EVENLY OVER THE SEMANTIC DOMAINS. The relative degree of analyzability can therefore be seen as a feature of the nominal lexicon as a whole, with differences found dominantly in the treatment of artifacts as items of acculturation on the one hand, and parts of the body on the other. There are languages where bodyparts are predominantly designated by morphologically simple lexemes on the one hand, but where the domain of artifacts, in contrast, is characterized by a high degree of analyzability, and vice versa. As will become clear in the following section, it is no surprise that artifact and body-part terms are the two semantic domains where no correlation in the overall analyzability is found, because there are areal differences in the distribution of analyzability in these domains.

5.3. ARE THERE AREAL FACTORS CONDITIONING THE DISTRIBUTION OF ANALYZABILITY?

It is conceivable that there are areal factors in play that govern the distribution of the prevalence of morphological complexity in the world's languages. When eyeballing the map plotting the world-wide distribution of the degree of morphologically complex terms in figure 1, some areal differences are apparent. For instance, there is a more or less contiguous area of low morphological complexity linking Eurasia (which also has some notable outliers such as Abzakh Adyghe, Ket, and Sora, discussion of which is in § 5.4.2.12.5.) with the North-Eastern part of Africa. Genealogically, it is interestingly the Afro-Asiatic languages in the sample, members of a family which is distributed over both Africa and Southwestern Eurasia through the Semitic branch, that pattern with Eurasia. Southeast Asia and Oceania appear to be areas of a moderate degree of morphological complexity overall, though there is some variation in particular in the New Guinea area. Marked differences again emerge in the Americas. In general, within the Americas, there is an West-East cline with respect to the variable, with lower values found in the Eastern part of North America. Likewise, in South America, languages of the greater Amazon region tend to score higher than languages spoken further in the West, in particular those spoken in an Andean environment. However, to really assess areality, mere eyeballing of maps is a dubious procedure (Cysouw 2005, Bickel and Nichols 2009), and statistical analysis is needed.

5.3.1. MORPHOLOGICAL COMPLEXITY, ALL DOMAINS

First, examining the percentages of analyzable terms in the whole set of 160 meanings under investigation, without recognition of differences that may exist with respect to the semantic domains investigated. there is no clear effect of area on the degree of analyzability using Dryer-6 (χ^2 = 10.1461, df = 5, p = .0712 by a Kruskal-Wallis rank sum test; all p-values in the further discussion of this section were obtained using this test). Under this breakdown, North American languages score very high. The corresponding plot is seen in figure 5.

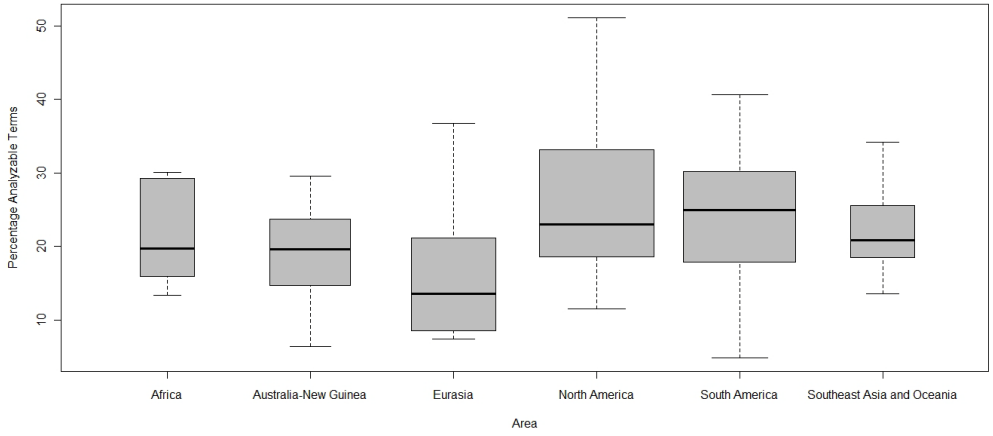


fig. 5: Areal breakdown of the degree of overall analyzability, using Dryer's (1992) breakdown

With the most fine-grained Nichols-11 breakdown, the differences between areas is closer to significance (χ^2 = 17.8067, df = 10, p = .05831). What this plot shows (and what is also suggested by impressionistically eyeballing the map) is that languages of Eastern North America score very high. Figure 6 plots the results.

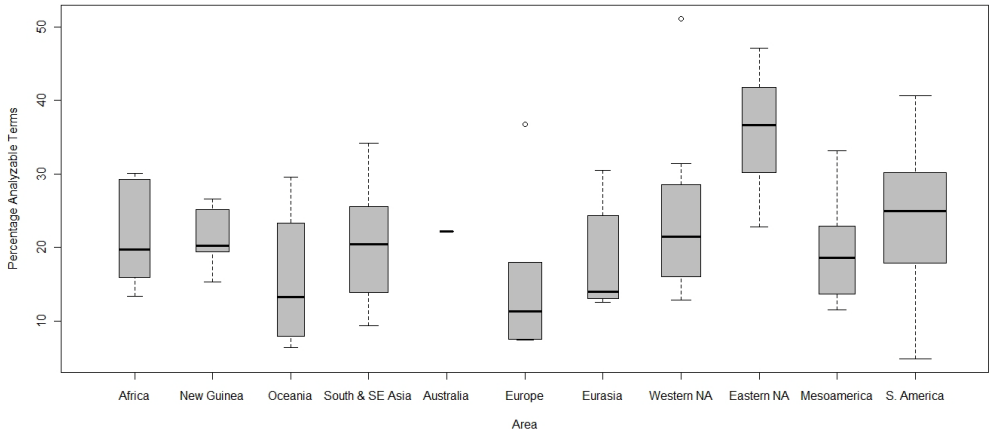


fig. 6: Areal breakdown of the degree of overall analyzability, using Nichols's (1992: 25-26) breakdown

An even stronger statistically significant difference emerges when using the broadest of the three partitionings: moving from the Old World via the Pacific into the New World, the

degree of overall analyzability rises significantly (χ^2 = 9.6076, df = 2, p = .008199), as visualized in figure 7.

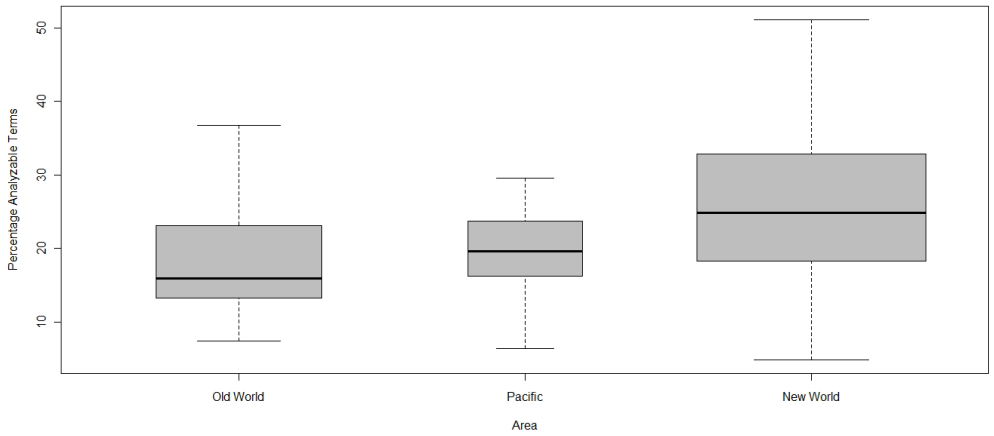


fig. 7: Areal breakdown of the degree of overall analyzability, using Nichols's (1992: 27) breakdown.

While this is an interesting result, it also raises a question, namely whether the difference is due to historical contingency rather than to language-inherent properties, as most of the modern-day artifacts were invented in the Old World and are relatively recent newcomers in many parts of the World. Consequently, it would make sense to expect that Old World-terms for artifacts are often unanalyzable due to their age, whereas neologisms in the New World have a clearly discernible morphological structure, as the artifacts have only be known for a short time span. Therefore, the same tests were performed, but with removing data from the artifact category from the data pool. If there remains a correlation, this would be an indication of genuine macro-areality. Under these testing conditions, the tendency for areality when using the Dryer-6 and Nichols-11 breakdowns ceases (χ^2 = 9.0625, df = 5, p = .1066 and χ^2 = 12.4319, df = 10, p = .2572 respectively). Although the same basic difference between the Americas on the one hand and other areas of the world, in particular Eurasia remains, this difference is not significant.

The statistical correlation with the three broadest possible sample areas as used in the Nichols-3 breakdown is weakened to p = .0336 (χ^2 = 6.7867, df = 2), with the ranking in the degree of analyzability from the Old World via the Pacific to the New World remains intact, as seen in figure 8. Thus, when artifact terms are not taken into consideration, there is NO CLEAR AREAL EFFECT ON THE DISTRIBUTION OF OVERALL MORPHOLOGICAL COMPLEXITY IN THE INVESTIGATED SLICE OF THE LEXICON UNDER THE SPLIT-UPS USED FOR TESTING. This should not necessarily be taken to entail that there cannot be areality on a smaller scale (cf. Bright 2004, tentatively also Nichols and Nichols 2007); to assess these, however, a much larger sample size would be needed. The findings should also not be interpreted in the sense that the semantic structure of analyzable terms as well as in colexification is not sensitive to lan-

guage contact and thus to areal effects (see extensive discussion in § 6.4.3). When it comes to sheer quantity of analyzable terms, however, areal factors appear to play an at best subordinate role.

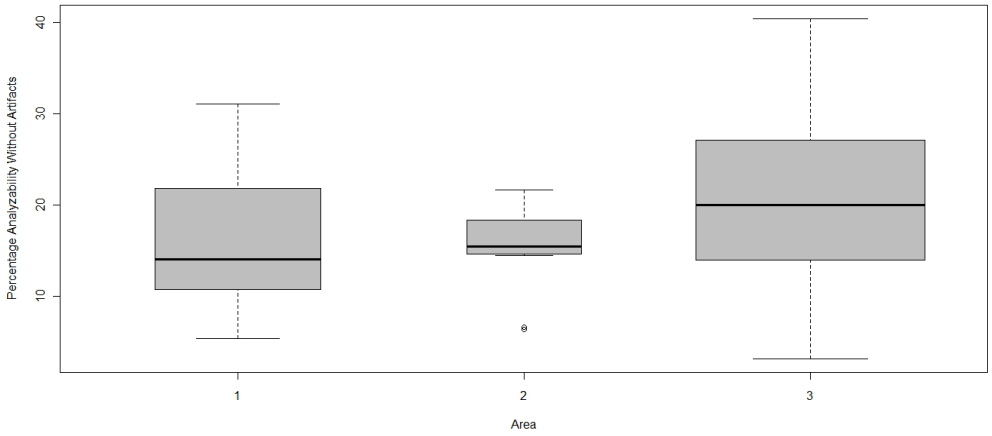


fig. 8: Areal breakdown of the degree of overall analyzability with artifacts removed, using Nichols's (1992: 27) breakdown.

Given that even large linguistic areas, if they exist, are the outcome of language contact, this is an indication that there seems to be little pressure on languages in contact to adjust the morphological structure of their lexicon (preponderance for complex terms in general on the one hand or preponderance of simplex lexical items on the other) to each other. Of course, to reiterate, this does not entail that calquing of morphologically complex expressions for a given referent does not occur – it does, but on a large scale, at the level of the lexicon at large, such tendencies seem to be rather weak.

5.3.2. Individual Semantic Domains

This section assesses differences in analyzability in the four semantic domains used in this study, with the same three breakdowns used for testing. There is no appreciable difference in analyzability of nature-related and topological terms under all three breakdowns (Dryer-6: $\chi^2 = 7.3432$, df = 5, $p = .1963$, Nichols-11: $\chi^2 = 7.1813$, df = 10, $p = .7082$, Nichols-3: $\chi^2 = 2.634$, df = 2, $p = .2679$). In contrast, as one would expect from the exercise of removing the artifact domain from the global calculations above, there is an areal effect on the analyzability of artifact terms under the Nichols-11 and Nichols-3 breakdown ($\chi^2 = 19.108$, df = 10, $p = .03891$ and $\chi^2 = 9.3812$, df = 2, $p = .009181$ respectively). The associated plots are seen in figures 9 and 10.

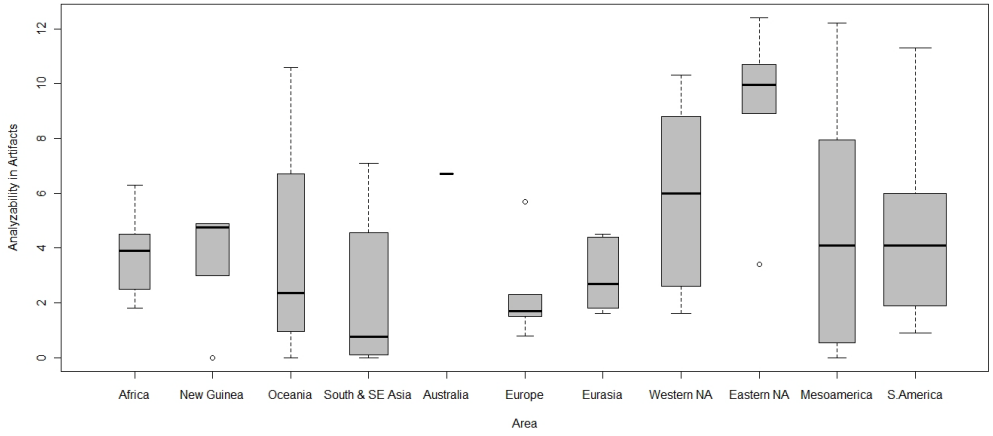


fig. 9: Areal breakdown of the degree of overall analyzability in artifacts, using Nichols's (1992: 25-26) breakdown.

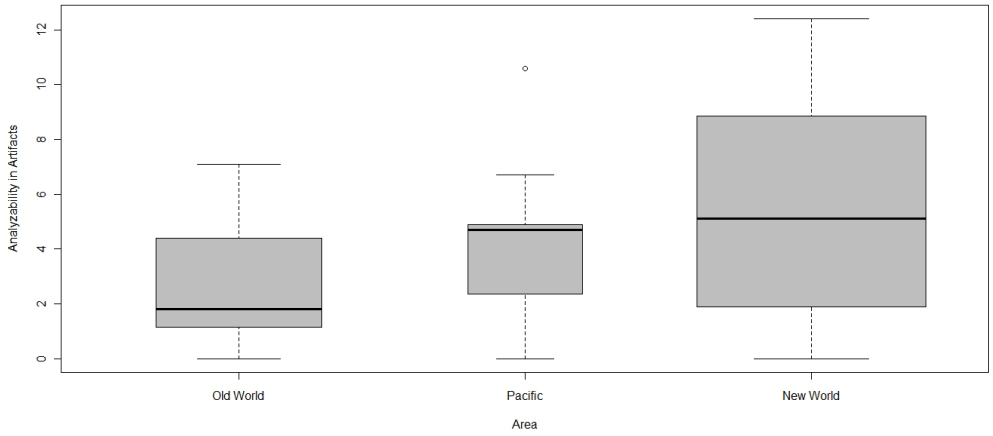


fig. 10: Areal breakdown of the degree of overall analyzability in artifacts, using Nichols's (1992: 27) breakdown.

The same tendencies – higher degrees of analyzability in artifacts in the Americas and very high degree of analyzability in artifact terms in North America – discerned by the application of the aforementioned breakdowns emerges when testing for Dryer-6, although the result is not quite significant (χ^2 = 10.0677, df = 5, p = .07334). A simple and straightforward conclusion follows, although it is hardly suprising: ANALYZABLE TERMS FOR ARTIFACTS ARE FOUND AT A HIGHER RATE IN THOSE AREAS OF THE WORLD WHERE THEY ARE RECENT ITEMS OF ACCULTURATION, and this notwithstanding the fact that another obvious option for lexical acculturation is borrowing of a word for a novel artifact from a contact language (this is

further discussed in § 5.4.2.7.1.). This obviously is an instance of what Haiman (1985: 149) calls the "iron horse" effect: "Languages tend to have complex periphrastic means of expressing notions that are unfamilar."

Moving on to the domain of body parts and body fluids, again there are differences between the areas tested under the Dryer-6 breakdown (χ^2 = 12.5949 df = 5, p = .02749). As seen in the corresponding plot in figure 11, it is South American languages that on average have the highest degree of analyzability in this domain.

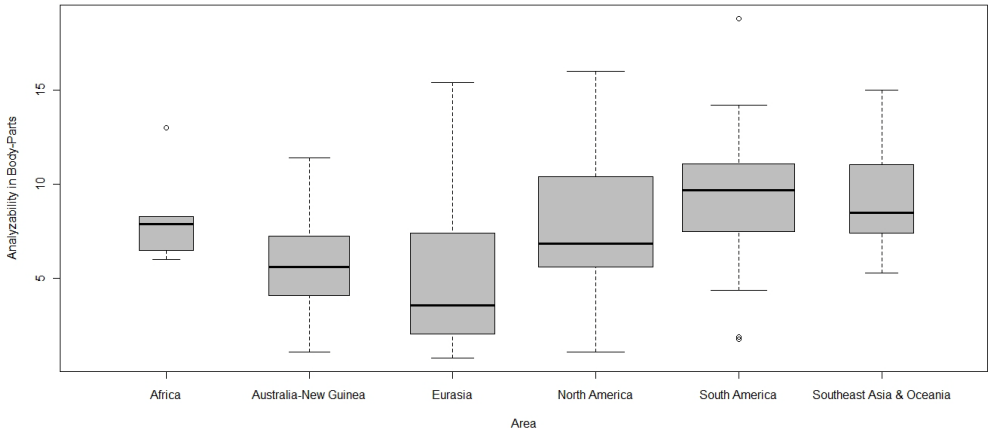


fig. 11: Areal breakdown of the degree of overall analyzability in body-part and body-fluid terms, using Dryer's (1992) breakdown.

The sharpest contrast is that between Eurasia, where body-part terms are least frequently analyzable and the Americas, in particular South America, where they are on average most commonly so. This is mirrored by the results of applying the Nichols-3 breakdown: as for the global values, there is a cline of rising degrees of analyzability moving from the Old World via the Pacific into the New World. The test for the Nichols-11 and Nichols-3 breakdowns are, however, not significant statistically (χ^2 = 15.0744, df = 10, p = .1294 and χ^2 = 4.9069, df = 2, p = .086 respectively), but also in the former, the high degree of analyzable body-part terms in South America is noticeable.

The difference between South America and the rest of the world in the evaluation based on Dryer-6 is mild, and may be due to two unrelated factors, namely the common process of derivation of body-part terms via sortal classifiers in a number of languages of the broader Amazon region (see § 4.4.1. for details), as well as a general increased presence of analyzable body-part and body-fluid terms, most of which are not particularly uncommon in the rest of the world in their semantic structure. Examples of body-part and body-fluid terms involving a sortal classifier from Bora (see Seifart 2005 for discussion of these in the Miraña dialect) are in (1.).

(1.) a. *nijpá-yu* urine-CL.ROUND 'bladder'
b. *máátyo-u* crying-CL.ROUND 'tear'

As for other complex terms in the domain of body-part terms, some languages of South America are unusual in that they have analyzable terms for 'mouth' (e.g. Tsafiki *fi'quí foró* 'language opening/hole') and 'stomach' (see Appendix E, 124 and 138). Also common are complex terms for 'vein,' most often via a metaphor involving either 'way, road,' such as Huambisa *numpa jinti* 'blood way' or sometimes on the basis of 'liana' (the conceptualization via 'way, road' is also heard of in other regions of the world).

There is no discernible areal effect when testing the domain of phases of the day and miscellanea (Dryer-6: $\chi^2 = 4.1783$, df = 5, $p = .524$, Nichols-11: $\chi^2 = 10.3555$, df = 10, $p = .4099$, Nichols-3: $\chi^2 = .8169$, df = 2, $p = .6647$) – an unsurprising result, given the heterogenous nature of this group of vocabulary items. Worth noting in this context is also the absence of areal effects on the overall degree of nature-related and topological terms, because also this group of meanings is fairly heterogenous. While the meanings clearly can be subsumed under a common denominator, it is still the case that they may be broken down into several smaller subdomains, such as the conceptualization of bodies of water, of things that have to do with fire, the heavenly bodies, parts and products of animals, etc. However, it is far from clear whether they form a lexical field that has the same degree of conceptual coherence that the domains of artifacts and body-parts possess. The results thus open up the possibility that it might be well-circumscribed semantic domains such as the body-part vocabulary and artifacts (demonstrated in Cognitive Psychology by priming experiments e.g. by Neely 1977 and Moss et al. 1995, in the case of artifacts assisted by historical contingencies), rather than the lexicon in general, which are likely to host areal clusterings of morphologically complex terms.

5.3.3. SUMMARY

Summing up, in the assessment of possible areality in the overall degree of morphological complexity, a statistical trend for languages in certain areas can be noted that, however, is so mild that one cannot discern a clear areal effect. Closer inspection of the individual semantic domains under investigation revealed that the degree of analyzability in artifacts and to a lesser degree in body-part and body-fluid terms is unequal in different areas of the world. Importantly, these differences mirror the general trend when evaluating overall vocabulary – rising degree of analyzability when moving from the Old World to the New World, in particular (parts of) North America. In effect, it appears that the structure of the vocabulary for body-parts and artifacts is at large responsible for the trend that is observable on a global scale, while the domains of nature-related and topological terms and phases of the day and miscellanea weaken it.

5.4. ANALYZABILITY IN THE LEXICON: TYPOLOGICAL PERSPECTIVES

5.4.1. STRUCTURAL CORRELATIONS TO THE DERIVED-LEXICAL CONTINUUM?

In § 4.7., a correlation between a preponderance of derived terms and the elaborateness of verbal person marking was established. Taking up this thread, this section explores whether there are further structural features that correlate with this distinction as to the type of analyzable lexical items. Since it is hitherto at large unclear what, if any, further factors may be relevant here, correlation tests using the data for the features in the World Atlas of Language Structures (Haspelmath et al. 2005) were performed. These tests are meant to be hypothesis-generating rather than hypothesis-testing. For that reason tests were carried out for the entire set of WALS features, regardless of how unlikely a connection between a given feature and the distinction between complex lexical items of the derived and lexical type may seem.

A word of caution in the interpretation of the findings is in order. While all languages in the statistics sample of the present study are also featured in WALS, it is not necessarily the case that very many datapoints are coded for them. While for Basque, for instance, a value is coded for 127 out of 138 features, a value for a meager eight features is available for Berik. In other words, it is the case for many features that the datapoints available for statistical testing are greatly reduced due to lack of coding in WALS (or the grammatical descriptions such coding presupposes), and in turn, the reliability of any statistical test depends to some extent on the available amount of coded data. Thus, the search for typological correlations on the basis of WALS need to be regarded as preliminary, in particular where the empirical database is small (see Wohlgemuth 2009: 187-189 for similar discussion). The preliminary tests on WALS yielded significant correlations with as many as ten WALS features:

(i) Imperative-Hortative systems (Van der Auwera et al. 2005):
$\chi^2 = 7.4559$, df = 3, $p = .0587$, Kruskal-Wallis Rank Sum Test
(ii) Order of Subject, Object, and Verb (Dryer 2005g):
$\chi^2 = 13.6505$, df = 5, $p = .01799$, Kruskal-Wallis Rank Sum Test
(iii) Order of Subject and Verb (Dryer 2005f):
$\chi^2 = 9.8122$, df = 2, $p = .007401$, Kruskal-Wallis Rank Sum Test
(iv) Order of Object and Verb (Dryer 2005d):
$\chi^2 = 5.598$, df = 2, $p = .06087$, Kruskal-Wallis Rank Sum Test
(v) Order of Adjective and Noun (Dryer 2005b):
$\chi^2 = 9.6764$, df = 2, $p = .007921$, Kruskal-Wallis Rank Sum Test
(vi) Position of Polar Question Particles (Dryer 2005i):
$\chi^2 = 8.3482$, df = 4, $p = .07963$, Kruskal-Wallis Rank Sum Test
(vii) Position of Interrogative Phrases in Content Questions (Dryer 2005h):
$\chi^2 = 8.2179$, df = 2, $p = .01642$, Kruskal Wallis Rank Sum Test
(viii) Relationship Between the Order of Object and Verb and the Order of Adjective and Noun (Dryer 2005j):
$\chi^2 = 11.4475$, df =4, $p = .02197$, Kruskal Wallis Rank Sum Test
(ix) Verbal Person Marking (see § 4.7)

(x) Nonperiphrastic Causative Constructions (Song 2005):
$\chi^2 = 5.5664$, df = 2, p =.06184, Kruskal Wallis Rank Sum Test

There are many features pertaining to word-order typology that yield significant or near-significant p-values, one is the overall classical Greenbergian word order typology, among the others are those looking at the order of subject and verb, the order of object and verb, and the order or adjective and noun specifically. However, as is well known, word order patterns are subject to areal pressure; in fact, the example of basic word order was the very trigger in linguistic typology to recognize that areal factors need to be taken into account when searching for universals in the classical sense (Dryer 1992). It is thus especially imperative to control for areal factors in the final analysis, using a Linear Mixed Effects Model (see § 4.7. for details), with the hypothesis to be tested in each case being that there indeed is a genuine influence of the above features on the degree of analyzability. Six of the above features (next to verbal person marking already discussed in chapter 4) survived closer scrutiny when controlling for areal effects[2] by employing the Mixed Model design familiar by now (in all models, the percentage of derived terms was square-transformed; for the feature concerning the order of subject and verb, no model could be built because even after various transformations residuals were still not normally distributed and the resulting model was therefore not valid):

(i) Order of Subject, Object and Verb (Dryer 2005g): p =.0247
(ii) Order of Object and Verb (Dryer 2005d): p =.0298
(iii) Order of Adjective and Noun (Dryer 2005b): p =.0053
(iv) Position of Interrogative Phrases in Content Questions (Dryer 2005h): p =.0018
(v) Relationship between the Order of Object and Verb and the Order of Adjective and Noun (Dryer 2005j): p =.0174
(vi) Nonperiphrastic Causative Constructions (Song 2005): p =.04003

The boxplot in figure 12 shows the distribution of the sampled languages with regard to the percentage of derived terms depending on the possible orders of subject, object and verb in the original statistics sample.

[2] Note that in principle the reverse situation, namely that effects only become visible rather than disappear when taking into account areal factors, is also conceivable. Since models have not been constructed for each of the WALS features, it is possible that there are some undetected WALS features for which a genuine correlation might exist.

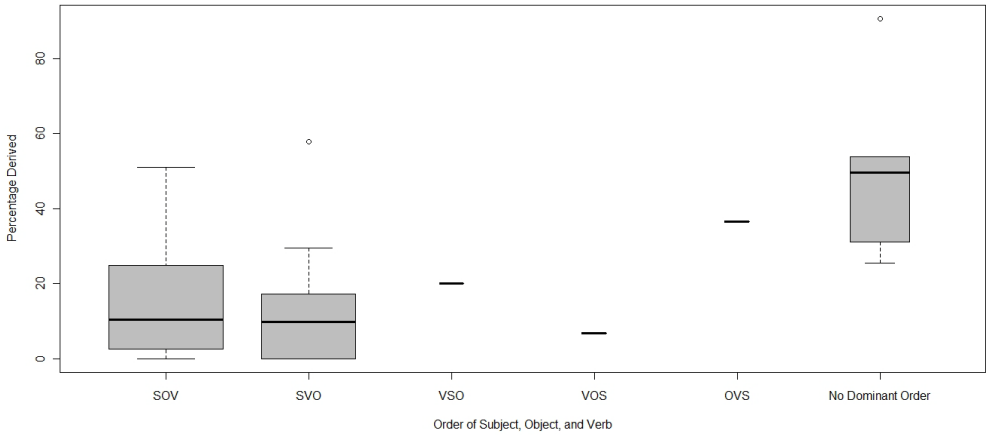


fig. 12: Percentage of derived terms depending on word order typology

As figure 12 shows, the main difference is not between languages with a fixed preferred word order of some kind, but rather between these and languages in which no particular grammatically conditioned word order is dominant.

The same basic observation can be made for the order of object and verb: here, too, it is the languages without dominant order that stand out in featuring an elevated number of derived terms, as seen in figure 13.

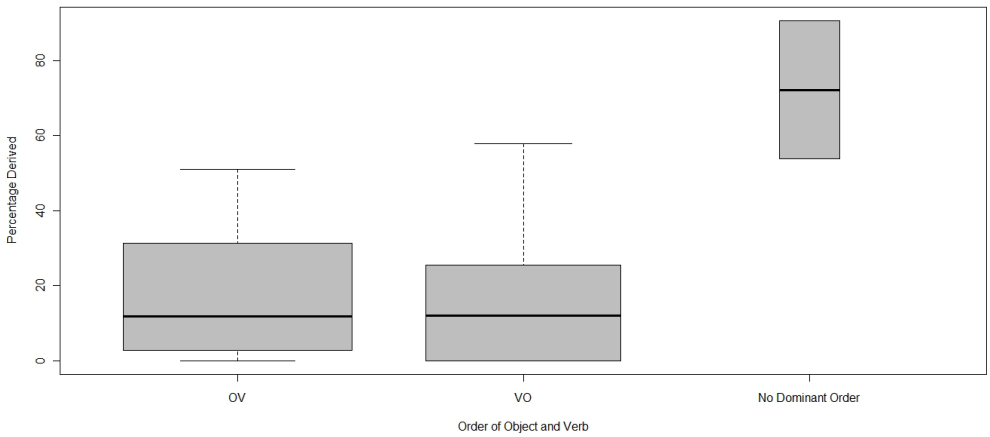


fig. 13: Percentage of derived terms depending on the order of object and verb

With respect to the order of adjective and noun, again it is the language where the order of these elements is not fixed that score highest with regard to the percentage of derived terms (figure 14).

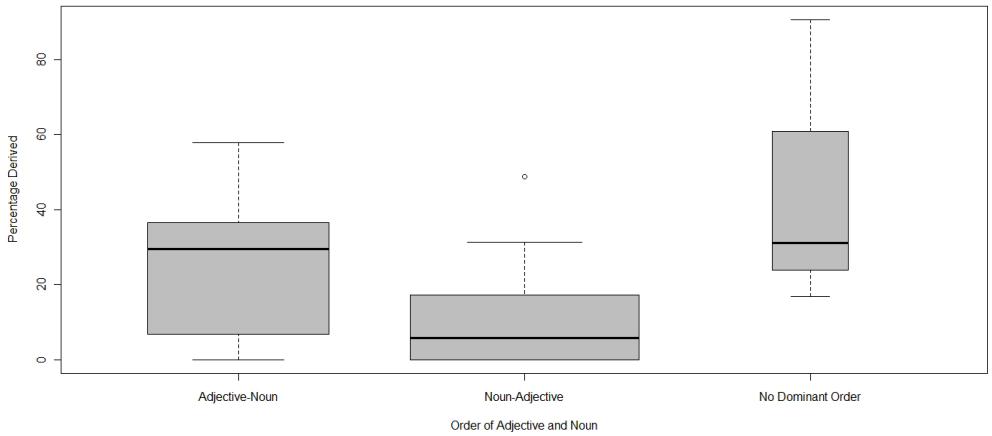


fig. 14: Percentage of derived terms depending on the order of adjective and noun

Bearing in mind the significant correlation with verbal person marking established in § 4.7, the trends seen so far seem easily accountable for: if information as to the arguments is coded morphosyntactically on the verb, there is functionally little need for fixed word order to make clear who does what to whom. Further, the correlation would also be additional evidence for a particular typological profile favoring terms of the derived type with synthetic morphology and concomitantly free word order.

Moving on to other significant correlations, as the boxplot in figure 15 shows, there is a drop in the percentage of derived terms in languages with non-initial interrogative phrases as opposed to those with initial interrogative phrases.

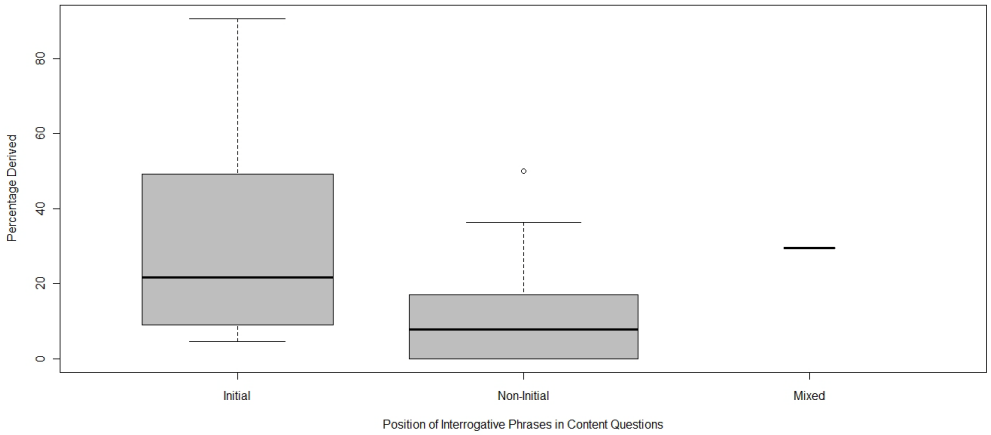


fig. 15: Percentage of derived terms depending on the position of interrogative phrases in content questions

Another significant correlation that is also independent of the basic word order typology according to Dryer (2005j) is that concerned with the order of object and verb on the one hand and that of the order of adjective and noun on the other. Notably, both variables on their own yielded significant interactions, as has already been discussed. Consistent with the findings made there, it is also here the languages grouped in the category showing an "other" behavior than the four logically possible main types (figure 16), that is, in Dryer's (2005j) coding, such languages where either or both order of object and verb or adjective and noun is not fixed or where constructions modifying nouns with adjectives are absent.

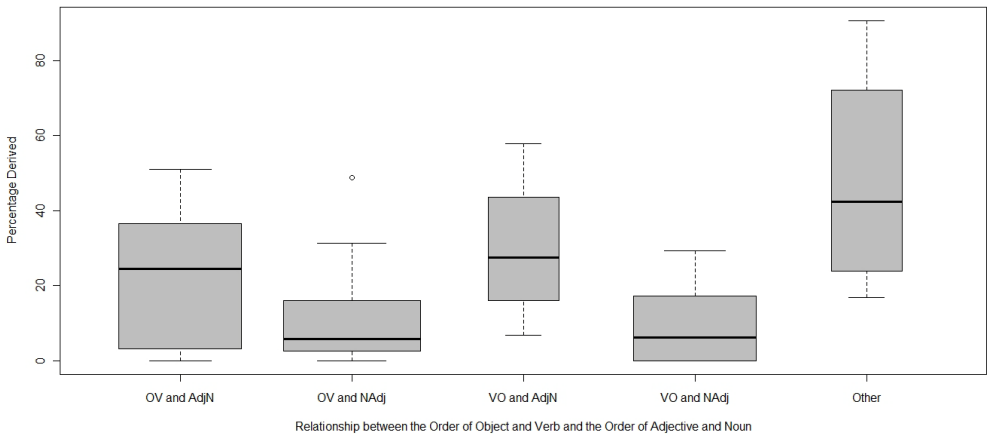


fig. 16: Percentage of derived terms depending on the order of object and verb and the order of adjective and noun

However, there still is one issue: ultimately, the initial tests leading to each of the hypotheses were part of a very large series of exploratory tests on the entire WALS dataset (not corrected for multiple comparison, as suggested for exploratory investigations by Bender and Lange 2001). Given the fact that for each test, there is a chance of $\alpha = .05$ percent that a significant result is obtained in the absence of any real effect, one can expect a number of about 7 tests with spurious significance simply due to chance. Therefore, it is furthermore imperative to cross-validate the results. Fortunately, this is possible, since data for more languages than those in the statistics sample were collected. From the remaining languages for which data is available for 65 percent or more of the investigated meanings, a genealogically balanced VALIDATION SAMPLE, as alluded to in § 3.3., was constructed. This includes the following languages, chosen randomly if more than one option was available for a particular language family: Swahili, Kanuri, Dongolese Nubian, Burarra, Kosarek Yale, Greek, Japanese, Vietnamese, Blackfoot, Comanche, Kashaya, Wintu, Yuki, Tuscarora, San Lucas Quiaviní Zapotec, Huambisa, Wayampi, Tsafiki, Ancash Quechua, Kapingamarangi, Mandarin, and Lesser Antillean Creole French. If a correlation is genuine, one should be able to replicate the results on data from entirely different languages, and hence also on those in the validation sample. Mixed Models were constructed for all of the features in the above list of six features. The estimates for the fixed effects were compared with those from the original models, and the correlation was taken to be genuine if they are within the range of the original estimate ± its standard error.

As for the exceptional behaviour of languages without fixed word order in the overall typology of order of subject, object, and verb, the estimate for the original sample is 3.7132 ± .9669, while that for the validation sample is only .2735, thus showing the same positive direction, but much more mildly and not within the range defined by the standard error around the estimate of the original sample. Hence, the effect must be rejected. The same is true for the subtypology looking only at the order of object and verb: the estimate from the original sample is 4.5910 and the standard error 1.7430, while the estimate for the validation sample is only .02062. For the correlation with the order of adjective and noun, validation is not possible because there is no language without a dominant order in the validation sample (the drop in the percentage of derived terms in languages with noun-adjective order present in the original sample, at any rate, cannot be replicated: original estimate is -1.9677 ± .7093 as opposed to -.2866 in the validation sample).

The correlation that can be most clearly replicated is the one which is at the same time most difficult to give reasons for, namely the position of interrogative phrases in content questions. The estimate for the difference in derived terms between languages with non-initial interrogative phrases and those with initial interrogative phrases from the original sample is -2.4418 ± .7494, while that of the validation sample is -1.8751, thus within the limits defined by the standard error (again, no evaluation of the behavior of languages with mixed position is possible since this group is very small and there are no representatives of it in the validation sample). Why this is the case is unclear; Dryer (2005h) does not mention correlations of this variable with other properties pertaining to word order, so that this feature seems unlikely to be a side-effect of a more easily explainable property.

With regard to the feature looking at the order of object and verb in relation to the order of adjective and noun, most estimates can be roughly replicated but notably not the most interesting one, namely the rise in derived terms in languages with a relationship other than the four major typological groupings recognized (estimates: -1.2492 ± .8800 vs. -0.39828, 1.1473 ± 1.3118 vs. -.08885, -1.8195 ± .9276 vs. -.42206, but 2.6139 ± 1.3118 vs. .57090).

The last of the significant correlations listed above, that with periphrastic causative constructions, is clearly disconfirmed by the evidence of the validation sample, at least for the group represented in both samples, namely morphological but no compound constructions (-1.298 vs. 4.86 ± 1.741).

Taken together, the results are suggestive, but the evidence from the validation sample suggests that the effect of word order typology, in particular the effect of free as opposed to fixed word order, is overestimated in the original sample and cannot at present be accepted as valid, while less obvious parameters of word order appear to have a replicable effect. Thus, verbal person marking seems to be the clearest correlate to the derived-lexical continuum that can safely be identified and at the same time explained functionally at present (which, of course, does not entail that it is the only one). Although the results are relatively meagre, the section at least serves to introduce the step-wise procedure used here to arrive at reliable correlations, and it will be made use of again in the following section, which approaches the question as to structural correlations to the degree of analyzability itself.

5.4.2. OVERALL MORPHOLOGICAL COMPLEXITY

5.4.2.1. Preliminary tests on the basis of WALS

This section seeks to elaborate on possible correlations between the degree of morphological complexity in the nominal lexicon as a whole and other typological properties of the sampled language, thus forming the major part of the entire chapter. The method employed is the same here as above: preliminary hypothesis-generating tests on the basis of WALS, elaborated on by more fine-grained analyses. Below are significant or near-significant correlations obtained by the preliminary tests.

(i) Consonant Inventories (Maddieson 2005a):
S = 10813.58, p = .01815, Spearman's ρ = -.391709
(ii) Consonant-Vowel Ratio (Maddieson 2005b):
χ^2 = 9.4684, df =4, p = .0504, Kruskal-Wallis Rank Sum Test
(iii) Syllable Structure (Maddieson 2005d)
S = 7627.729, p = .02406, Spearman's ρ = -.3980442
(iv) Possessive Classification (Nichols and Bickel 2005c):
S= 1026.556, p < .0001, Spearman's ρ = .6866435
(v) Semantic Distinctions of Evidentiality (de Haan 2005):
χ^2 = 9.8448, df = 2, p = .007282, Kruskal-Wallis Rank Sum Test
(vi) Order of Adjective and Noun (Dryer 2005b):
χ^2 = 6.5014, df = 2, p = .03875, Kruskal-Wallis Rank Sum Test

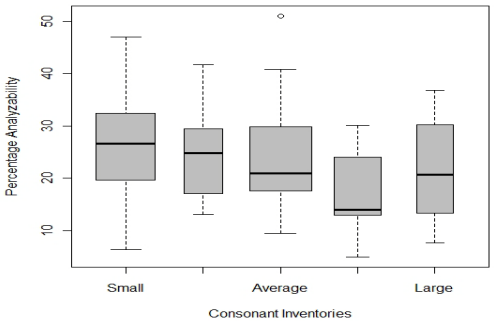
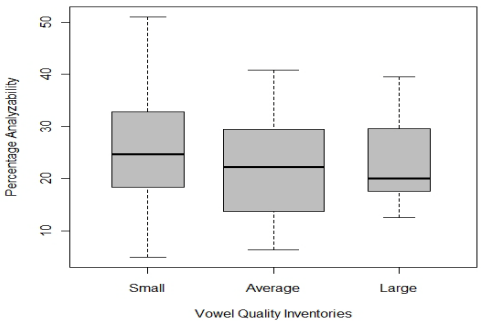
(vii) Order of Demonstrative and Noun (Dryer 2005c):
$\chi^2 = 8.8377$, df = 4, $p = .06529$, Kruskal-Wallis Rank Sum Test
(viii) Predicative Adjectives (Stassen 2005a)
$\chi^2 = 5.9285$, df = 2, $p = .0516$, Kruskal-Wallis Rank Sum Test
(ix) Purpose Clauses (Cristofaro 2005)
$\chi^2 = 3.0855$, df = 1, $p = .079$, Kruskal-Wallis Rank Sum Test

5.4.2.2. Elaborating on the preliminary findings with regard to phonology

Surprisingly, two phonological features are tested positively for significant interaction on the basis of the WALS data, and another one yields borderline significance. Apparently, the smaller the consonant inventory and the simpler the structure of the maximal syllable, the higher the amount of morphologically complex lexical items will be. Given that two features in the area of phonology yield significance, some real interaction is likely to go on between phonology and lexicon.

However, as noted above, results need to be interpreted with caution at this stage since there are many gaps in the data. In order to arrive at reliable results, and to examine whether the correlation can be substantiated, the policy adopted here is to amend the WALS database with data from published materials for the relevant phonological features in order to fill gaps in cases where the statistical testing on the basis of the WALS data revealed significance and the correlation appeared to be amenable to meaningful interpretation. In doing so, additional data were also gathered for the other pertinent feature in this area, namely vowel quality inventories, although here only a very weak negative correlation (Spearman's $\rho \approx -.08$) that is clearly not significant statistically ($p = .6368$) was found when testing on the limited WALS data. Information from published materials was coded in precisely the same fashion as in the relevant WALS features (Maddieson 2005a, b, d, h) to ensure compatibility of the data; furthermore, phonemes indicated to be non-native and restricted to loanwords were not counted in making coding decisions. Data are in appendix C. A problem for analysis is that the phonological features are highly unevenly distributed areally as revealed by Kruskal-Wallis rank sum tests (and as also suggested by Maddieson 2005a, d, h).

It is thus again particularly important to control for areal factors in the final analysis, using Linear Mixed Effects Models. The findings on the basis of the enhanced datasets for phonology are seen in table 1.

Feature	*p*-value	Plot
1. Consonant Inventories[3]	*p* =.0234 estimate: -1.977	
2. Vowel Quality Inventories	*p* = .5896, estimate : -.9965	

[3] An apparent clerical error in Maddieson (2005a) was corrected before performing analysis: Oneida, according to one of the sources consulted by Maddieson (Abbott 2000), should be, with nine distinctive consonant phonemes (/l/, /w/, /y/, /n/, /t/, /k/, /s/, /ʔ/, and /h/), coded as having a small, not moderately large consonant inventory. This coding decision would be valid even if one recognized voicing as distinctive in the alveolar stop as proposed in some analyses, but not followed by Abbott (2000).

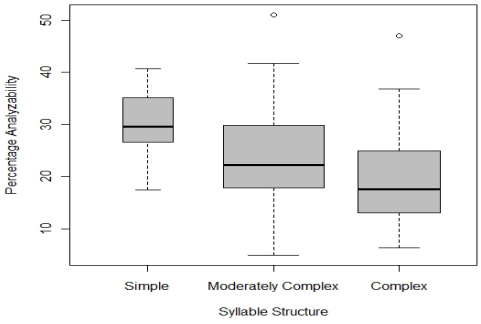
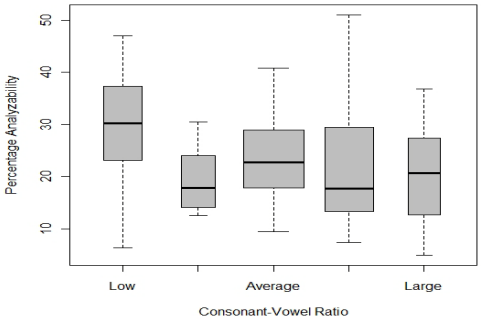
3. Syllable Structure[4]	p =.0102, estimates: -7.688, -13.053)	
4. Consonant-Vowel Ratio	p =.0401, estimate: -2.028	

table 1: Differences in the degree of analyzability depending on phonological factors, computed on an extended dataset based on WALS

Testing on the amended datasets substantiates the interaction between consonant inventories (the correlation is now a little weaker here, but the distribution is much more even and observations in each group sufficiently large) and syllable structure, and also confirms the insignificance of vowel inventory size on the degree of analyzability.[5] A correlation that is a bit difficult to interpret is that regarding the consonant-vowel ratio. This is calculated by simply dividing the number of distinctive consonants by the number of distinc-

[4] Maddieson's (2005d) coding decision was revised with respect to Tetun, which is coded by him as having a complex syllable structure, presumably because of Morris (1984) mentioning weakly articulated excrescent consonants in the syllable onset in emphatic speech. In spite of this, Tetun was coded as having a moderately complex syllable structure given Van Engelenhoven and Williams-van Klinken's (2005) description of Tetun syllable structure as (C)V(C).

[5] Note that also for this feature, it is true that languages with smaller inventories tend to have more analyzable terms. However, unlike for the other features, there are areal factors in play: when not controlling for area, a borderline significance emerges also for this feature. When areal differences are taken into account in the Mixed Model, significance ceases, so that it is not a valid cross-linguistic generalization to say that there is a direct influence of vowel inventory size on analyzability in the lexicon. This example underscores the importance of taking into account areal biases when formulating cross-linguistic generalizations.

tive vowel qualities. Ultimately, this entails that a languages with both few consonant and vowel phonemes and languages with both very many consonant and vowel phonemes will end up receiving similar scores in the ratio of consonant to vowels, and thus, this measure is in principle no measure of phonological complexity per se. However, it is important that the variance within consonant inventories is much greater than that within the vowel inventory system: while the number of distinctive consonants in Maddieson's (2005a) sample ranges from six to 122, the number of distinctive vowel qualities varies only between two and fourteen (Maddieson 2005h).[6] An effect of this is that, as noted by Maddieson (2005b), languages with large consonant inventories typically also have a large consonant-vowel ratio. Thus the areal distribution of the figure for consonant-vowel ratio sometimes overlaps with that for consonant inventories. This is noticeable for instance in the American Northwest. Many languages spoken in this region have both large consonant inventories and a high consonant-vowel ratio, whereas in Eastern South America, many languages have small consonant inventories and also a low consonant-vowel ratio. This at first glance somewhat hidden dependency is likely the key to explain why a significant correlation between the consonant-vowel ratio and the degree of analyzability is found.

Cross-validating the results using the validation sample already used above after amending data also for the languages in this sample (see appendix C for data), it turns out that the estimate for Consonant Inventories as a predictor in the validation model is -.1140, thus within the limits of that for the statistics model ± its standard error (-1.9768 ± .8416), and also well within these limits for the estimates for syllable structure (-4.292, compare -7.688 ± 3.486 and -12.952, compare -13.053 ± 3.990 respectively) and for the Vowel-Consonant-Ratio (-2.678, compare -2.0282 ± .9556). Hence, all correlations appear genuine.

To sum up, the smaller the consonant inventory of a language, the simpler the maximal syllable (and the lower the consonant-vowel ratio), in short, THE SIMPLER THE PHONOLOGICAL SYSTEM, THE MORE COMPLEX THE NOMINAL LEXICON CAN BE EXPECTED TO BE. As suggested by Maddieson (2005d: 55), there is some evidence that syllable structure complexity and consonant inventory size are interrelated cross-linguistically. This issue is discussed in more detail in § 5.4.2.8.

The correlations on a global scale already at this point help to explain some variation in particular areas of the world. For instance, it is common knowledge that there are significant differences in the size of consonant inventories in North America, to the effect that languages in the western part typically have larger inventories when compared with those in the east (Sherzer 1973: 774, Mithun 1999: 15).[7] Using the Rocky Mountains as a watershed dividing western from eastern languages, this difference turns out to be mirrored in the degree of morphologically complex lexical items as shown in figure 17, and

[6] Though note that variables such as length, nasalization and diphthongs are largely discarded in Maddieson's (2005h) coding scheme in order to make the data more readily comparable, and this approach is followed here for consistency.

[7] However, in North America, there is also "increasing head-marking as opposed to dependent marking going from west to east" (Fortescue 1998: 80). See § 5.4.2.12.5. for discussion of this as a possible factor.

this fact explains to some extent the areal hotspot of languages with a highly analyzable nominal lexicon detected in § 5.3.[8]

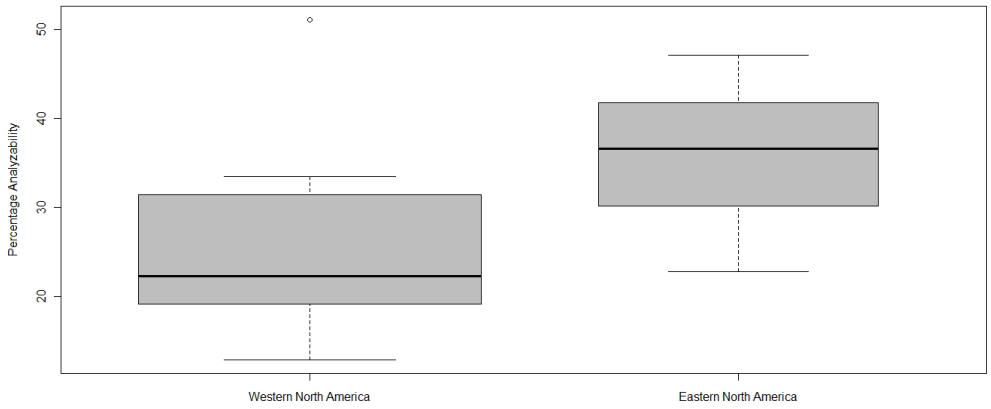


fig. 17: Differences in the percentage of analyzable terms between Western and Eastern North American languages

However, it is important to point out that these correlations are a statistical generalization, and there are languages which behave unexpectedly. In other words, there is no law in the sense of a classic implicational universal that a simple phonological system will in all cases trigger a lexicon characterized by morphological complexity. The most extreme case of a language that goes against the trend in the sample is Buin. Like its unrelated (or unrelatable) neighbor Rotokas, which is also spoken on the island of Bougainville, Buin has a very small phoneme inventory and simple (C)V syllable structure, and yet the degree of analyzable lexical items is one of the lowest both in the larger New Guinea area and worldwide (but see § 5.4.2.6. for a possible explanation of the behavior of Buin).

While establishment of a cross-linguistic correlation is of value in itself and in a way is more solid than proposed explanations for the correlation (Dryer 2003), it is important to note that skewed distributions are not an explanation in themselves, but rather something that needs to be explained (Cysouw 2003: 99), be it by appealing to functional, cognitive, or other factors. As a first step to get to the bottom of the correlations with phonological properties, and also in the light of concerns as uttered by Plank (2003: 138) that typology should not merely be an exercise in statistics, in the following sections three case studies will demonstrate apparent influences of phonological factors on complexity in the lexicon in synchrony and changes in diachrony in greater detail. Polynesian languages, Mandarin Chinese, and the "Papuan" language Vanimo will serve as examples. Particularly interesting is the case of Mandarin Chinese, for which there is actual diachronic evidence for the development of a largely compound-based lexicon and its phono-

[8] The outlier in Western North America is Kiliwa, which, although spoken in the west, has like other Yuman languages an average-sized consonant inventory as opposed to the large systems more common in the west.

logical motivations. For the case of Polynesian languages there are striking diachronic developments in the phonology that suggest a similar line of argumentation. Even though the earlier stages of Polynesian are not attested, they are at least fairly well reconstructed. The case studies will furthermore also serve to bring to light other aspects of phonological structure and its repercussions on the structure of the lexicon which can then be elaborated on.

5.4.2.3. Case Studies

5.4.2.3.1. *Case study I: Polynesian.* The Polynesian languages are a low-level branch of the Oceanic subfamily, which is in turn one of the best-established subgroups of the Malayo-Polynesian languages, themselves one of the primary branches of Austronesian. The Polynesian languages for which data were sampled (Hawaiian, Samoan, and Kapingamarangi) consistently score quite high with respect to the degree of analyzable terms, higher than most non-Polynesian Austronesian languages in the sample. At the same time, they are known to have very small phoneme, in particular consonant inventories. Since it is known that ancestral Proto-Oceanic had a considerably larger number of phonemes, and since the historical reconstruction of developments in that subgroup is in a fairly advanced stage, it should be possible to trace the developments in the Polynesian lexicon historically, departing from the Proto-Oceanic stage. Table 2 charts the Proto-Oceanic sound system as given by Ross (1988: 93).

	velarised bilabial	bilabial	alveolar	palatal	velar	postvelar
stop	bw	p b	t d	c j	k g	q
trill			r dr			
sibilant			s			
nasal	mw	m	n	ñ	ŋ	
liquid			l			ʀ
glide	w			y		

table 2: Proto-Oceanic consonant inventory, from Ross (1988: 93)

This amounts to a number of 23 consonant phonemes. Syllable structure was probably already fairly simple at this stage and is posited to be (C)V "with the option of a word-final consonant" (Lynch et al. 2002a: 66) which is lost in most daughter languages. Notably, loss of word-medial consonant clusters that were permitted in Proto-Malayo-Polynesian is one of the features that defines Proto-Oceanic as a subgroup (Lynch et al. 2002: 66).

Further down the genealogical tree, one finds the so-called Proto-Central-Pacific subgroup of Oceanic, believed to have been spoken between 100-800 BC on the Fiji islands (Trudgill 2004: 308, table 1, compiled from various sources; for a more general overview of the history of the Austronesian expansion see e.g. Pawley 1999). The Proto-Central-Pacific phase must have been relatively brief since evidence in terms of shared innovations is sparse, and it is thought to have been a dialect chain rather than a homogenous language (Pawley 1996b: 390; 2009: 529fn7). In any case, this dialect chain gave rise to both the Polynesian languages as well as Fijian and Rotuman, although the precise relationship of the latter to the other daughter languages is somewhat unclear. In other words, Proto-

Polynesian is a primary branch of Proto-Central-Pacific. The Proto-Central-Pacific consonant system (from Geraghty 1986: 290) is charted in table 3.

	bilabial	dental	alveolar liquids	alveolar fricatives	palatal	velar	labiovelar	glottal
fricatives	v			c	z	x		
stops	p	t	r			k	kw	ʔ
prenasalised obstruents	b	d	dr	s	j	q	qw	
nasals	m	n	l		ñ	g	gw	
glides	w				y			

table 3: Proto-Central-Pacific consonant inventory, from Geraghty (1986: 290).

Here one encounters 25 consonantal proto-phonemes, a little more than Proto-Oceanic had (Pawley, as cited in Trudgill 2004: 310, believes that the number of distinct segments was somewhat lower, "around 21"). Differences to the Proto-Oceanic situation include the presence of a contrastive series of prenasalized obstruents, a series of labiovelars and phonemic glottal stop.

Significant phonological simplification sets in on the way from Proto-Central-Pacific to Proto-Polynesian. Developments include (data from Geraghty 1986, see also Pawley 1996b: 392-393):

(i) merger of proto-phonemes */p/ and */b/, */d/ and */t/, */dr/ and */r/, */k/ and */q/, as well as */k/ and */kw/. In short, prenasalization is lost as a distinctive feature; these developments are shared with Rotuman. */r/ further apparently began to merge with */l/ in Proto-Polynesian under unclear conditions, a change that was completed in Proto-Nuclear-Polynesian. Proto-Fijian retains the majority of these contrasts.

(ii) Proto-Central-Polynesian */z/ changes to */h/ in Proto-Polynesian, in some instances the reflex is also *s*, i.e. a partical merger.

(iii) Merger of */j/ with */s/, */t/ or */d/; */j/ is only retained in Rotuman

(iv) Loss of */y/

(v) Merger of */ñ/ with */n/

(vi) Merger of labiovelars: */k/, */kw/, */q/, */qw/ fall together in */k/, and */g/ and */gw/ in */g/. Fijian retains the contrast between labiovelars and velars.

(vii) Merger of */x/ with */ʔ/

These developments leave Proto-Polynesian, most likely spoken between 500 BC and 200 AD on the Fiji islands (Trudgill 2004: 308, table 1), with a system of thirteen consonant phonemes (Biggs 1978): stops */p/, */t/, */k/ and */ʔ/, fricatives */f/, */s/ and */h/, nasals */m/, */n/ and */ŋ/, as well as */w/, */l/ and */r/.

Subsequently, Proto-Polynesian split into what is being called Proto-Nuclear-Polynesian, the common ancestor of the sampled languages Hawaiian, Samoan and

Kapingamarangi, on the one hand and Proto-Tongic on the other. As a result of this development, Proto-Nuclear Polynesian lost two further distinctive consonants inherited from Proto-Polynesian, namely */r/ and */h/, leaving it at eleven consonant phonemes. Proto-Central Eastern Polynesian, an even more direct ancestor of Hawaiian, additionally lost phonemic glottal stop, and, finally, Hawaiian itself is distinguished from its direct ancestor by merging nasals */n/ and */ŋ/ as well as */f/ and */h/, leading to its present-day system of eight consonant phonemes (glottal stop is reintroduced in Hawaiian by regular change of alveolar stops). The resulting Hawaiian phonological system allows for the generation of only 162 distinct syllables (Maddieson 1984: 22); it is seen in table 4.

	Bilabial	Dental-alveolar	Alveolar	Velar	Glottal
stops	p			k	ʔ
liquids			l		
fricatives					h
nasals	m	n			
glides	w				

table 4: Hawaiian consonant inventory, adapted from Biggs (1978: 708), Elbert (1979: 10-13)

Samoan also lost the inherited phonemic glottal stop just to reintroduce it as the regular reflex of */k/, but else maintains the Proto-Nuclear-Polynesian system, yielding the ten consonant phonemes seen in table 5.

	Bilabial	Labio-dental	Lamino-alveolar	Dorso-Palatal/ Dorso-Velar	Glottal
stops	p		t[9]	(k)	ʔ
liquids			(r), l		
fricatives		f, v	s		(h)
nasals	m		n	ŋ	
glides					

table 5: Samoan consonant inventory, adapted from Mosel and Hovdhaugen (1992: 20-21)

Samoan is characterized by pervasive diglossia. The above inventory is that of the tautala lelei; note that /k/ and /r/ are restricted to loanwords, and /h/ to loanwords and a few native interjections. The tautala leaga has three phonemes less, due to merger of /t/ and /k/, /n/ and /ŋ/, and /r/ and /l/. Mosel and Hovdhaugen also note that /p/ and /f/ are interchangeable for many speakers.

In Kapingamarangi, both */f/ and */s/ become /h/ by regular sound change, and the number of distinctive consonants in the language is thus nine, as seen in table 6.

[9] May "be pronounced as an apico-dental, apico-alveolar, lamino-dental, or lamino-alveolar stop." (Mosel and Hovdhaugen 1992: 20). /n/ may also be articulated as a lamino-dental or apico-alveolar (Mosel and Hovdhaugen 1992: 21).

	Bilabial	Dental	Alveolar	Velar	Glottal
stops	p	t		k	
liquids			l		
fricatives					h
nasals	m	n		ŋ	
glides	w				

table 6: Kapingamarangi consonant inventory, adapted from Biggs (1978: 708)[10]

Present-day Fijian, in contrast, has nineteen consonant phonemes (the Boumaa dialect in addition has phonemic glottal stop, Dixon 1988: 12). Table 7 shows the inventory.

	Bilabial	Labio-dental	Apico-dental	Apico-alveolar	Dorso-velar	Glottal
stops	p		t		k	ʔ
prenasalized stops	b		d		q	
liquids				r, dr, l		
fricatives	v	f	c	s		
affricates				j		
nasals	m		n		g	
glides				y	w	

table 7: Fijian consonant inventory (Boumaa dialect), adapted from Dixon (1988: 13)

In summary, "these unusually small inventories are simply the phonological end point of a millennia-long reduction in the number of consonants as languages spread further and further into the Pacific" (Trudgill 2004: 310).

Rensch (2002: 191) discusses these diachronic phonological developments, and states that, in connection with the simple syllable structure, "[t]he result is a high number of homonyms," which are fed in addition by syntagmatic phonological changes (for instance, to adduce an example from the present study, Hawaiian, Kapingamarangi, and Samoan *lā, laa*, and *lā* all colexify 'sun' with 'sail.' The terms were distinct in Proto-Polynesian, having the shape **laʻaa* and **laa* respectively according to Elbert and Pukui 1986: 188 and collapsed due to elision of intervocalic glottal stop). At the same time, Rensch relates these observations to statements in the literature as to "language inherent therapeutic devices which prevent or heal harmful clashes," to which he takes the Polynesian evidence to be a counterexample. However, it appears that an increase in segmental length of lexemes in Polynesian languages, and, on account of the evidence of this study, a substantial number of which by means of formation of complex lexemes, took place in

[10] Note that Lieber and Dikepa's (1974: 375) brief description of Kapingamarangi phonetics and phonology consistently distinguishes between slightly and heavily aspirated versions of each consonantal segment.

Polynesian languages, and this can be construed as just such a therapeutic device, a line of thought that will be discussed in much greater detail in following sections.[11]

In fact, the difference in morphologically complex lexical items in Polynesian when compared to the remaining Austronesian languages in the sample is consistently higher (figure 18).

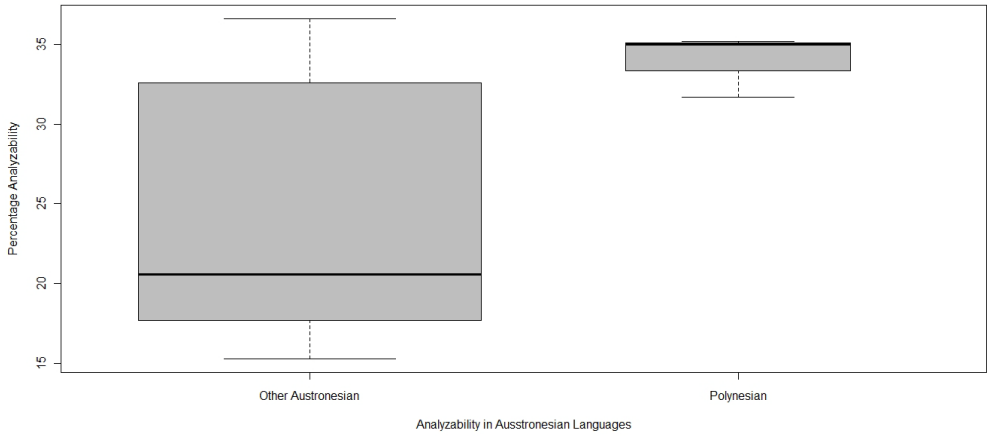


fig. 18: differences in the percentage of analyzable terms between the sampled Polynesian and other Austronesian languages

As becomes clear from the plot in figure 18, there is wide variation in the non-Polynesian languages, and some of the languages score as high as Polynesian. Notably, however, for those languages with high scores outside the Polynesian subgroup, similar accounts in terms of phonology are available: Tetun, for instance, independently has developed a small inventory of consonants phonemes (thirteen, according to van Engelenhoven and Williams-van Klinken 2005: 737, table 26.1) and concomitantly a lexicon that is characterized by a high degree of analyzability.

However, there are some unexpected results with respect to Austronesian that appear to run counter to the proposed account for the relatively high degree of analyzability in the lexicon of Polynesian languages. Fijian in fact does not do what one would

[11] Trudgill (2004), with reference to Polynesian specifically, argues that severe simplification of phonological systems is more tolerable in tightly-knit, isolated communities due to a high amount of shared information and cultural knowledge that can be pressuposed (while also noting that phonological systems of languages spoken by societies with the above characteristics may alternatively also be unusually large). Thus, in the case of Polynesian specifically, these societal factors, together with the absence of language contact with concomitant second-language acquisition did not prevent the development of very small phoneme inventories. In response to Trudgill, Pericliev (2004) tests the hypothesis of size of speech community and size of phoneme inventory on a large scale, with negative results, and Hajek (2004) argues that, at least in languages of New Guinea and the Pacific area, areal diffusion is apparently the most prominent factor responsible for the reduction of phonological inventories. For the purpose of the correlation between simple phonological systems and morphological complexity in the lexicon, one can remain agnostic as to what societal factors, if any, caused the significant decrease in inventory size in Polynesian, and simply note that this reduction did happen.

expect under the hypothesis. While it has a similarly simple syllable structure as Polynesian and a similar basic five-vowel system, its consonant inventory is far from being as drastically shrunk as the one of its Polynesian kin, but still the language shows a relatively high degree of analyzability in the lexicon that is comparable to that of Polynesian languages. Given its comparably large system of consonants, one would expect homonymy to be less of a problem in this language; however, Dixon (1988: 237), writing on the Boumaa dialect of Fijian specifically, informs that "[t]here is a good deal of homonymy in Fijian," and if this is indeed the case for whatever reason (e.g. a low functional load of some phonemes), then the same explanation of the structure of the lexicon is available for this case as well, in particular because the facts concerning syllable structure and vowel system fit the overall picture. Conversely, Rotuman, also a close congenitor of Polynesian, receives a low score in analyzability. It is at present unclear whether this is due to the phonemic inventory not being reduced as drastically as in Polynesian languages (with fourteen consonants, see table 8) or due to multiple layers of loanwords from a wide variety of sources (Biggs 1965, Schmidt 2003).

	Labial	Dental/Alveolar	Palatal	Velar	Glottal
stops	p	t		k	ʔ
liquids		r, l			
Affricate			tʃ		
fricatives	f, v	s			h
nasals	m	n		ŋ	
glides					

table 8: Rotuman consonant inventory, adapted from Vamarasi (2002: 7)

At present a convincing account of these facts is missing. As noted above, the data make a statistically significant cross-linguistic generalization possible, but it is far from being an absolute universal on a global scale, so counterexamples are neither unexpected nor damaging to the overall correlation. Thus, in spite of Fijian and Rotuman not quite fitting into the picture with respect to reduction of the consonant inventory, the difference between Polynesian languages and the other Austronesian languages in the sample is clearly present, and this difference is accountable in the way outlined above, the somewhat problematic case of Fijian notwithstanding.

Pawley (2009) primarily investigates retention rates in a variety of Oceanic languages spoken on the Solomon islands on the basis of a list of 60 vocabulary items, but also reports (2009: 529fn7) very high retention rates of basic vocabulary in both Proto-Central-Pacific (60 out of 60) and Proto-Polynesian (54 out of 60), which are therefore quite conservative Oceanic languages in terms of vocabulary replacement. Thus, in a time span of approximately 1,000 years after the breakup of Proto-Oceanic (though see below for potential problems with dating), a very large percentage of vocabulary items is retained in Proto-Polynesian. In contrast, percentages of retained vocabulary computed for purposes of glottochronology, as those in Elbert (1953) and Biggs (1978), indicate that among themselves, Polynesian languages have on average about 50 per cent shared vocabulary (as a remainder, such statements pertain to "basic" vocabulary as defined by the Swadesh list or similar lists), the highest figure being 72% shared vocabulary between Tongan and East

Uvea and the lowest 33% between Tongan, Samoan, and Tahitian. Thus, after the breakup of the still lexically conservative Proto-Polynesian, vocabulary replacement appears to have accelerated to a certain degree, and this may be a possible effect of the creation of morphologically complex neologisms replacing inherited vocabulary. However, "all the Fijian languages and some Polynesian languages (especially Tongan)" are considered lexically conservative when compared with some other Oceanic languages (Pawley and Ross 1995: 61), which speaks against replacement of inherited vocabulary on a larger scale.

On the other hand, one could also construct an argument in favor of relatively rapid vocabulary replacement out of the available data. The purpose of Pawley (1996b) is to defend from the point of view of linguistics the traditional view that posits a pause of around 1,000 years in the settlement of Eastern Polynesia after the settlement of Western Polynesia against claims by Irwin (1992, and other publications), who instead argues for more or less continuous settlement without major breaks. The linguistic correlate of that time span is the development of Proto-Polynesian out of Proto-Central-Pacific, and Pre-Polynesian is a term adopted by Pawley (1996b) to refer to the time before the breakup of Proto-Polynesian. Relying on glottochronological dates, Pawley (1996b: 400) notes that "[t]o allow only 400-500 years for the Pre Polynesian period would be to suppose a rate of lexical change over this period probably unparalleled in the subsequent history of any of the 30 individual Polynesian languages." Certainly, Pawley's argumentation is stringent, and this is not the place to contest archaeological evidence; however, it seems worth noting that, at least for the development of the lexical profile of Polynesian, the marked decrease in distinctive consonants on the way from Proto-Central-Pacific to Proto-Polynesian, which is continued in the Polynesian daughter languages, but which was already very advanced at the time of the breakup of Proto-Polynesian, may have accelerated lexical change in an unusually fast manner. There is no evidence for any other language present in the Fiji-Polynesia area at the time of Proto-Central-Pacific, and thus no indication that such accelerated rates of lexical change could be contact-induced (Pawley 1996b: 395). Within Polynesian, Pawley (1996b: 399), evaluating the data from Biggs (1978), states that "[t]he apparently more innovative languages include Samoan, Tahitian, Kapingamarangi and Nukuoro," two of which figure in the present sample. This again is compatible with the hypothesis of an increased rate of vocabulary replacement by coinage of complex terms because of limited expressive possibilities resulting from the shrunk consonant inventory, although the differential rates of vocabulary replacement, under the present account, still beg for a conclusive explanation, given that all Polynesian languages have experienced severe phonological simplification. Of course, any statements about the degree of vocabulary retention on the one hand and the degree of morphological complexity on the other are a function of the meanings selected for investigation, and if apparently conflicting results emerge, this may be attributable to the difference in vocabulary items that are investigated.

Summing up, the Polynesian case study is not entirely conclusive, and there are loose threads emerging from it that cannot be woven together into a coherent and conclusive account here. But the statistical difference between the degree of analyzablity in Polynesian when compared with other Austronesian languages remains a fact, as does the

heavy phonological simplification these languages have undergone since the time of Proto-Oceanic.

5.4.2.3.2. *Case study II: Mandarin Chinese*. There is one case where a temporal coincidence between phonological simplification and an increase in morphologically complex lexemes (compounds, in this case) is well-established and where the developments have been traced historically in a variety of publications: Mandarin Chinese. Typically, however, this process is discussed in phonological terms as disyllabification of the Mandarin lexicon, but, as will be seen in the following discussion, word-formation plays a major role in bringing about this pervasive change. Still, it is necessary to carefully distinguish between the phonological and morphological facts in the discussion (Feng 1997).

The basic facts concerning the simplification of Chinese phonology are as follows (dates from Arcodia 2007 throughout unless attributed to another author): According to Feng (1998: 213), syllables of CCVCC structure were possible in Old Chinese (ca. 1200 BC – 300 AD). Arcodia (2007: 84) and Feng (1998: 224) also mention clusters of up to three segments in both onset and coda in Old Chinese as spoken around 1000 BC. In Middle Chinese, in contrast, the syllable structure was simplified to CV(C) around 800 AD, with the additional constraint that only a subset of the available consonants, three nasals and three stops, were allowed in coda position. In Mandarin, only nasals appear in the syllable coda, initially three, later only two (Lin 2001: 84). In addition, whereas in Middle Chinese 35 distinct consonants could be found in onset position, only 20 are allowed for in Mandarin, and voicing was lost as a distinctive feature in consonants (Shi 2002: 73). Furthermore, affixation was lost. For instance, Old Chinese suffix *-s gives rise to a suprasegmental feature (tone) in Middle Chinese (Haudricourt 1954). Summing up, cluster simplification, loss of affixation, and reduction of possible consonants in syllable coda occurred. "As a result of consonant-cluster simplification, the number of phonologically distinct syllables in the language decreased dramatically" (Feng 1997: 213).

With the phonological simplification ongoing, the process of disyllabification of the lexicon set in. While it is true that disyllables are attested already in Old Chinese, it is equally true that they were relatively rare and that their number increased exponentially only at a later point of time. Text counts performed by Shi (2002: 75) suggest that the process of disyllabification (using disyllabic verbs as examples) reached its peak in the period between the 5th and 12th century AD; text counts by Feng (1997: 219) suggest an earlier date, to the effect that the process of disyllabification was "undergoing relatively large scale development during and after the Han dynasty" (Packard 2000: 265), that is between the 2nd century BC and the 2nd century AD. What are the precise mechanisms to disyllabify the lexicon that can be detected? Shi (2002: 76) lists the following:

(i) suffixation (Shi 2002: 74 mentions the nominal suffixes *-zi*, *-er*, and *-tou*)
(ii) "monosyllabic words are juxtaposed with synonyms," i.e. the creation of semantically redundant complex lexemes
(iii) replacement of inherited monosyllabic words by new disyllabic ones

(iv) reduplication
(v) conventionalization of adjacent syntactic constituents in discourse as fixed expressions that enter the lexicon, cf. Feng (1997: 208-209).

These facts make clear that the process of disyllabification is largely brought about by standard mechanisms of word formation (with a broad definition of word formation as employed for present purposes that does not exclude syntactic mechanisms from this category as long as they serve to form fixed expressions that enter the lexicon). Feng (1997) provides a number of enlightening examples, comparing a Classical Chinese text by Mencius (born around 370 BC), with a later commentary on the same text by Zhao Qi, written around 200 AD, that is, in the time of the Han dynasty in which disyllabification is said to have set in.

(2.) a. Mencius
shengren qie you guo
sage-person also have mistake
'Even sages make mistakes'

b. Zhao Qi
shengren qie you miu-wu
sage-person also have false-mistake
'Even sages make mistakes' (Feng 1997: 205)

(3.) a. Mencius
Wang Liang tianxia zhi jian gong ye
Wang Liang world 's lousy artisan PRT
'Wang Liang is the lousiest artisan in the whole world.'

b. Zhao Qi
Wang Liang tianxia bi-jian zhi gong-shi ye
Wang Liang world clumsy-lousy 's artisan-artisan PRT
'Wang Liang is the lousiest artisan in the whole world.'
(Feng 1997: 214, slightly adapted)

There are competing accounts for the increase in the number of (morphologically complex) disyllables while at the same time the language underwent phonological simplification, most prominently the 'functional' and the 'phonological' (Packard 2000: 266). According to the functional account, as summarized by Packard (2000: 266), societal and economic growth and concomitant introduction of new ideas during the Han dynasty (which is likely to be the time in which disyllabification of the lexicon set in on a larger scale) led to an increased need to coin neologisms in order to fill the gaps in the lexicon as no words existed to designate them. Once the lexicon was saturated with newly coined compounds, phonological distinctions, under this account, were given up since they were no longer

needed to keep words distinct, a job that had been taken over by the increased word length due to compounding. Also, it is argued that the increase in compounds were created by the preference in Chinese tradition to have pairs of entities, a solution which Feng (1997: 219) finds "theoretically unattractive, and empirically problematic." In contrast, the so-called phonological hypothesis (note that there are terminological inconsistencies: what is being called the phonological hypothesis by others is called the functional hypothesis in Feng 1997) states that the developments occurred rather in the reverse order, and that the increase in disyllabic lexemes is a functional response to the reduced complexity of the phonological system. Thus the label 'phonological hypothesis' is somewhat misleading, since it is at its core functional as well, albeit language-internally. This explanation is mentioned frequently, and is most often evaluated positively (Packard 2000: 265-267, Shi 2002: 72-74, see also Li and Thompson 1981: 14, and further references in Shi 2002). In the words of Lin (2001: 10):

> The change that started out with syllable simplification did not stop at the production of homophones. Indeed, one should not normally expect one change in a language to have no further effect, as chain reactions are common in language evolution. In the case of Mandarin, it is at least partially due to the great number of homophones in the language that another significant historical development was effected – the disyllabification of words. Earlier, we mentioned that M[iddle] C[hinese] had predominantly single-syllable words. However, when the syllable simplification was producing a great number of homophones, the dialect had to make some adjustment to avoid ambiguity. One logical measure would be to enlarge the word in size, and that was exactly what happened. ... Disyllabification has not wiped out the monosyllabic homophones; it has merely demoted them from the level of the word to the level of the morpheme in the dialect (Lin 2001: 10)

In favor of this account, Shi (2002: 74) importantly points out that southern varieties of Chinese preserve more traits of the inherited phonological system of Old and Middle Chinese when compared with the northern ones (including Mandarin).[12] This correlates with the fact that often southern Chinese monosyllabic words correspond to disyllabic compounds (probably of the semantically redundant type, see below) in northern varieties. "A simple explanation is that [southern] Cantonese has more phonological devices to distinguish lexical forms and thus does not need as many disyllabic words" (Shi 2002: 74).

Packard (2000: 267), in discussing the merits and drawbacks of the two accounts, also favors the phonology-based account "because it involves two processes that remain operative in the modern language: the continued simplification of the Chinese phonological system ... and the continuation of 'compounding' as a way of forming new words." Feng (1997: 213), however, raises some doubts regarding this explanation since the functional load formerly carried by segmental phonemes was in part taken over by suprasegmental features. Instead, Feng argues that the development of compounding is due to disyllabic foot formation that was established in the time of the Han dynasty, and which is itself due to the loss of bimoraic feet already occuring in Old Chinese. The simplified syllables re-

[12] Also, Mandarin has one of the smallest numbers of tonal opposition of any of the varieties of Sinitic (Mian Yan 2006).

sulted in a decline of syllable weight, to the effect that one syllable alone could not form the minimal prosodic unit of the foot anymore (Feng 1997: 226). Under this account, then, the phonological process of disyllabification is initially in terms of its motivation independent of the increase in compounds at the morphological level. In other words, Feng's account is a more sophisticated version of the phonological explanation, since the prosodic structure was ultimately caused by simplified syllable structure, and the causing factor here, as well as in more traditional versions of the phonological account, is ultimately simplification in phonology. Although Feng explicitly argues against traditional phonological explanation, because of the problem he sees with counter-functional compounds, Packard (1997: 7) summarizes his position as being an "insightful adaptation" of the traditional phonological view.[13]

5.4.2.3.3. *Case study III: Vanimo, Papua New Guinea.* An intriguing case in the literature for a correlation between extreme phonological simplicity and complexity in the lexicon (with examples almost exclusively drawn from the nominal domain) is Vanimo, a New Guinea language of the Skou family as discussed by Ross (1980).[14]

The segmental phoneme inventory features eight vowels, all of which may occur nasalized and sometimes contrasting phonemically with their non-nasalized counterparts. There are thirteen consonants and nine allowed consonant clusters, of which two are doubtful; in addition, there are three phonemic tones. Syllable structure is (C)V, where C is a single consonant, or one of the abovementioned clusters. However, for Vanimo specifically, a very important additional factor that appears to constrain the structure of the lexicon even more heavily than the sheer phonological facts is that "[t]he syllable and the morpheme appear to be – or to have been until quite recently – coterminous." By multiplying 20 consonants and (secure) consonant clusters with 16 vowel qualities and three tones, Ross establishes "that the number of possible morphemes in Vanimo cannot exceed – or have exceeded – 960, an extraordinarily low number. Semantically these resources are in effect less, as each verb paradigm has five or six different morphemes" (Ross 1980: 101). Note that there appears to be a correlation of such a situation, in which the syllable and morpheme are coextensive, with the presence of tones, as seen in the discussion of Mandarin Chinese. Ross states that the lexical concomitant of the phonological simplicity is the "attribution of very wide meanings to some morphemes, and their combination of other morphemes which act as specifiers." In the nominal domain specifically, these combinatorics primarily result in noun-adjective and noun-noun compounds. Ross (1980: 102-105) provides ample examples for the operation of compounding to counter the scarce distinctiveness of the language's morphological resources. A selection of examples, with some adaptations to simplify accessibility, are in tables 9 and 10.

[13] Similarly but independently, Duanmu (1999) argues that metrical structure favors disyllabic words (which have been present to a smaller degree already in older stages of the language and have been introduced to some extent by newly coined neologisms) in syntactic non-head position, which also accounts for the frequent semantic redundancy of Mandarin compounds emphasized throughout by him.

[14] Note that Skou, the language of the eponymous family in the sample, is not part of the core sample due to insufficient data as defined in chapter 3.

Simplex: *paŋ* 'arm, wing, frond'	
Complex term	Modifier
paŋè 'arm'	*è* 'bone, long object'
dìŋpaŋ 'wing'	*dìŋ* 'bird'
èŋpaŋ 'coconut frond'	*èŋ* 'coconut'
ñéŋpaŋ 'snake'	*ñéŋ* 'octopus'
yípaŋ 'sago frond'	*yí* 'sago pudding, food'

table 9: Vanimo compounds based on *paŋ*, adapted from Ross (1980: 103)

Simplex: *boŋ* 'intangible substance'	
Complex term	Modifier
yaboŋ 'smell, odour'	*ya* 'thing'
téboŋ 'smoke'	*té* 'fire'
øboŋ 'dust'	*ø* 'ground, earth'
høboŋ 'fog'	*hø* '??'
èŋboŋ 'coconut milk'	*èŋ* 'coconut'

table 10: Vanimo compounds based on *boŋ*, adapted from Ross (1980: 103)

Interestingly, Ross (1980: 101) also notes that "[n]oun compounding of this kind appears to be an areal feature of the West Sepik coastal region;" it is at present not entirely clear whether, if this is indeed an areal feature, it is due to similar grammatical restrictions as found in Vanimo. While not spoken in that area, there is another New Guinea language in which a very similar situation in the nominal lexicon obtains, and for which the same explanation is available: Toaripi. For this and for the closely related Orokolo, Brown (1972: 157) notes: "Both T[oaripi] and O[rokolo] have many homonyms or near homonyms and it often becomes necessary to guard against confusion of meaning. A way of doing this employed by both [languages] is to use compound expressions in place of the simple nouns." Brown does not mention what may have given rise to the situation of exuberant homonymy, but it seems extremely likely that phonology is the responsible factor here as well. Toaripi has nine consonant and eight vowel phonemes (Brown 1972: 119-120), and the syllable structure that can be inferred from the aforementioned source appears to be maximally (C)V.

5.4.2.4. Tonality and Morphological Complexity

Preliminary tests based on the WALS data revealed no discernible interaction between the presence or absence and the nature of tonality with the measured degree of morphological complexity. However, two of the case studies, that of Mandarin Chinese and the Papuan language Vanimo, revealed that there is a potential connection between tonality and the structure of the lexicon, which is why data on tonality for the languages in the statistics sample was gathered as well.

When assessing whether there is an impact of tonality on analyzability in the nominal lexicon, the *p*-value for tone as a predictor when distinguishing between simple and complex tone systems reaches only a very weak borderline significance at $p = .1057$. As the plot in figure 19 shows, the clearest contrast is between tonal- and non-tonal lan-

guages in general, while the differences between languages with simple as opposed to complex tone systems is not dramatic. In fact, instead of a constant upward trend analogous to rising complexity in the tone system, the analyzability score drops as one moves from simple to complex tone systems, which casts some doubt on the effect of tonality on the degree of analyzability. On the other hand, when simplifying the distinction to a binary opposition of tonality vs. non-tonality, the difference turns out to be significant at $p = .0342$ (estimate: 5.611).

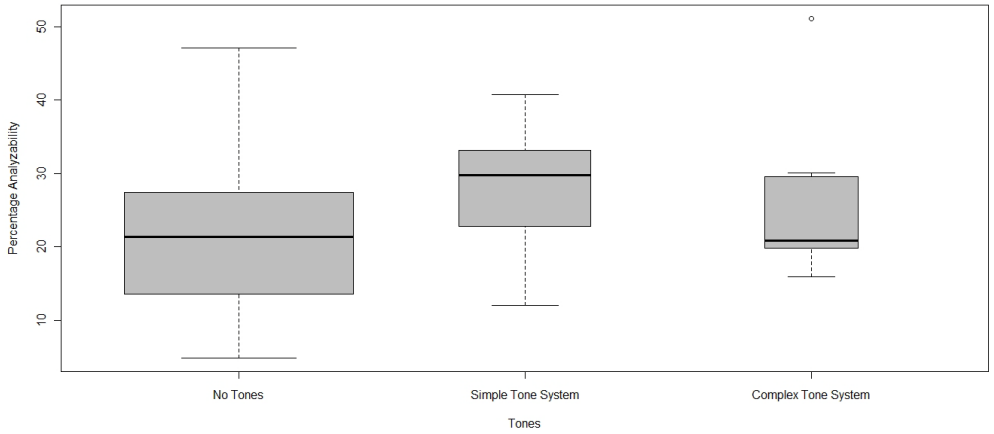


fig. 19: Correlation between tonal complexity and morphological complexity in the lexicon (data partly from Maddieson 2005e)

Thus, in spite of the unexpected non-linearity of the correlation, tonality can be added to the list of relevant phonological factors explaining morphological complexity: TONAL LANGUAGES TEND TO HAVE MORE ANALYZABLE NOMINAL LEXICONS THAN NON-TONAL ONES, an exception being Ket, a language with a relatively high degree of analyzability and thus unusual for Eurasia (see Vajda 2004b for an analysis that posits tones in Ket; however, the language is treated as non-tonal by Maddieson 2005e in the light of competing analyses).

On the one hand, the correlation between the presence of tones (whether the tone system is simple or complex) and an increased degree of analyzability in the lexicon is quite surprising in the light of Maddieson's (2005e) discussion of interrelations between tonality and other phonological properties. According to his data, increased tonal complexity typically goes hand in hand with a rise in the number of consonants as well as the number of distinctive vowel qualities (although he is also noting that the latter correlation in particular is subject to some areal variation, with a non-systematic relationship in particular in the Americas). In contrast, in the present study, one obtains a correlation between the degree of analyzability and tone as well as a correlation between this variable and consonant systems, which is quite surprising given that Maddieson's data indicate a correlation between tone and large segmental inventories! Maddieson (2005e) in particular suggests that a decrease in the complexity of the tone system goes hand in hand with a

decreasing number of languages with moderately complex syllable structure, while, as complexity in the tone system increases, the number of languages with complex syllable structure decreases.

On the other hand, the fact that there is a significant interaction between morphological complexity and tone is not too surprising, given that for instance the present-day tone system of Mandarin Chinese came into being as the complexity of syllables decreased. In fact, Mandarin Chinese is just one example for a broader scenario of tonogenesis outlined by Matisoff (1973). Without mentioning any particular reason why this should be so, Matisoff (1973: 77) states that as a prerequisite for the development of a full-fledged tone system "a language must have a basically *monosyllabic* structure (i.e. the morphemes must be only one syllable long)" and that "[t]here is something about the tightly structured nature of the syllable in monosyllabic languages which favors the shift in contrastive function from one phonological feature of the syllable to another" (Matisoff 1973: 28). This is in line with the observation made in the previous chapter as to the monosyllabicity of certain languages in Southeast Asia. Interestingly, one finds such a situation not only in the case study on Mandarin Chinese, but also in Vanimo and, in the Americas, for instance in Hupda (Epps 2008: 41 also notes "a strong preference for isomorphism between the morpheme and the syllable" in Hup, which has a two-way tonal contrast). According to this view, phonetic perturbations in the fundamental frequency of vowels due to neighboring consonants (see Hombert et al. 1979 for more phonetic details), which are an ordinary phonetic phenomenon, were phonemicized in Tibeto-Burman languages (which have a monosyllabic word structure) when phonological simplification broke down the originally complex phonological structure of the Tibeto-Burman monosyllables. In the words of Matisoff (1973: 79), "[i]t was only when the old consonantal system had decayed through cluster simplification, losses, mergers that the daughter languages were forced to exploit those pitch-differences for contrastive purposes." Importantly, in the highly abstract general scenario of tonogenesis as outlined by Matisoff (1973: 82-83), the impact of all these phonetic-phonological processes on the lexicon comes into play:

> Thus we may imagine a hypothetical language at Stage A: it is monosyllabic, but the number of possible syllables is very large, since there is a rich system of syllable initial and -final consonants. ... Different syllables have different pitches, but the language can afford to ignore this fact, since it is having no trouble keeping its utterances apart. [In stage B] its initial- and final-consonantal systems are breaking down. ... Homophony rears its ugly head. In desperation the language casts about for ways to protect its contrasts. Although each morpheme is still monosyllabic, the language now creates bisyllabic or even trisyllabic compounds in order to disambiguate homophones or near-homophones, so that the word is no longer monosyllabic. ... Meanwhile the number of vowels has increased and lexically contrastive tones have arisen, exploiting the previously redundant pitch-differences among syllables (emphases removed).

Matisoff (2001: 295) mentions that homophony is also notorious in the Loloish branch of Tibeto-Burman (compare also Bradley 2002: 1070), with compounding as a disambiguation strategy to counter it.[15]

Discussions of tonogenesis have a certain bias towards Southeast Asia, because the mechanism involved were first studied for languages of that area. However, there are also other possible diachronic paths leading to the emergence of phonemic tone. For instance, tonal contrasts in Cheyenne reflect Proto-Algonquian vowel length (Frantz 1972), with new lenght contrasts being introduced by the (sporadic, according to Goddard 1990: 104) loss of Proto-Algonquian *p and *k (Frantz 1972: 223). However, also in Cheyenne, the emergence of tonality goes hand in hand with at least some degree of segmental simplification, albeit of a different kind than for instance in Mandarin Chinese.

According to Ratliff (1992), a certain type of tone language (her Type A languages) can be defined by the fact that tone is used predominantly for contrastive lexical purposes, but only to a minor extent for morphological ones. Ratliff's example is White Hmong. This language has almost no segmental morphology, monosyllabic roots, a complex tone system, and a calculated number of 754 possible combinations of segmental contrasts without tonal contrasts factored in. According to Ratliff (1992: 135), "[s]ince syllables are usually coextensive with morphemes, almost all possible combinations need to be realized as morphemes. There is a high level of homophony as well," and thus, "[t]one must be used for lexical discrimination when there are not enough other resources avaliable in a tone language to do the job" (Ratliff 1992: 137). This statement is in agreement with Matisoff's diachronic scenario in which tone needs to be exploited to keep lexical morphemes distinct as phonological complexity decreases, next to an increase of the morphological complexity of words. Tone, as seen above, is suggested to be correlated cross-linguistically with monosyllabic words.

The correlations with phonological features are able to account for the behavior of many languages in the sample with respect to the degree of analyzability in their lexicon, but not all. For instance, Buin was already mentioned as an example of an "aberrant" language above. An entire region of the world where variation in analyzability cannot well be accounted for on the basis of the correlations so far established is the Caucasus. However, there is a way of accounting for this variation. This account is interrelated in a way with the relevance of the shape of the lexical morpheme for differences in analyzability suggested by the discussion of tone (although none of the Caucasian languages are usually described as being tonal, but see Kodzasov 1999, who argues that at least some Nakh-Daghestanian languages feature tone systems). Another reason to believe that this is a relevant factor comes from a number of languages with a relatively high degree of analyzability, tonal or non-tonal, for which authors note that lexical morphemes are normally monosyllabic, and that any elements departing from this shape in being longer can be identified diachronically as old compounds. This is the case for instance for Ket (Werner

[15] Matisoff (1973: 91n30) claims that "[i]nstances of this process abound in the world's languages. In some American English dialects where *pin* and *pen* are homophonous, the words are replaced by the compound forms 'stick-pin' /stɪkpɪn/ and 'ink-pen' /ŋkpɪn/, respectively."

1997: 46: "[h]istorisch lassen sich die meisten mehrsilbigen und auch manche einsilbige Wörter auf Komposita zurückführen" / "historically, most polysyllabic and also some monosyllabic words can be traced back to compounds") and Kiowa (Watkins 1984: 75: "there are polysyllabic nouns which can be tentatively regarded as old compounds on the basis of identification of at least one element with synchronically occurring forms. Still other polysyllabic nouns are entirely unanalyzable, but given the monosyllabic structure of roots and the tonal patterns of known compounds, they can safely be inferred to be old compounds").

But first, to make the argument more palpable and to show how it can account for variation that is otherwise not explainable, the following final case study presents the basic relevant facts about Caucasian languages.

5.4.2.5. Case Study iv: Variation in the Caucasus

There are three languages spoken in the Caucasus in the present sample, corresponding to the three major families that are indigenous in this region of the world: Abzakh Adyghe (Northwest Caucasian), Laz (Kartvelian), and Bezhta (Nakh-Daghestanian). These languages share a number of grammatical features, such as pervasiveness of ergative alignment. They also have some commonalities in the phonological systems, which typically feature a cross-linguistically unusual large number of consonant phonemes, to the effect that the Caucasus is sometimes said to form a linguistic area, although large-scale areality is disputable.[16] The languages are in addition all non-tonal (though again compare Kodzasov 1999 for a different point of view). Yet, there are also marked typological differences between the languages, and the Caucasus is also a region notable for its great linguistic diversity, both in terms of the large number of languages it hosts in a comparatively small territory as well as structural-typological variety (Comrie 2008). For instance, Northwest Caucasian and Kartvelian languages have many traits typically associated with polysynthesis, such as a rich system of verbal inflection. In contrast, morphological complexity is more pronounced in the inflection of nouns in Nakh-Daghestanian; particularly noteworthy are the rich case systems. There are sharp differences among the sampled Caucasian languages with respect to the degree of analyzable terms in the nominal lexicon that is presently investigated. Laz and Bezhta score very low and are thus typologically "normal" in the larger context of Eurasia, which is characterized by a comparatively low degree of analyzable terms when compared to the situation in the rest of the world (cf. § 5.3.). In contrast, Abzakh Adyghe is the language with the highest percentage of analyzable nouns in all of Eurasia.

The discussion in Rayfield (2002) makes clear that these differences can be accounted for by morphophonological factors. These factors, however, are less noticeable when examining the values assigned to the individual languages in the coding of their phonological properties. All are coded as having large consonant inventories and complex syllable structures. Bezhta and Laz have average-sized vowel inventories, while that of

[16] Tuite (1999), for instance, argues that the prevalence of ergativity in this region can equally well be explained by universal typological preferences, although not denying that the Caucasus has been a contact zone for a considerable amount of time.

Abzakh Adyghe is coded as being small. Rather, the structure of the lexicon, in particular the degree of analyzability, apparently has something to do in particular with restrictions on the phonotactic structure of the lexical root. As for the nominal lexicon of Kartvelian, according to Rayfield (2002: 1039), "the wide variety of syllable structures allow for a large number of non-homophonic roots, mono- and bi-syllabic" and the phonological inventories, together with the allowance for complex consonant clusters "give the language group enough resources to produce tens of thousands of distinct monosyllabic lexemes." Boeder (2005: 9-10) confirms the complexity of consonant clusters in Kartvelian languages, although noting that permittable clusters in Mingrelian and Laz are somewhat less complex than those of Georgian. Furthermore, in Kartvelian, there are marked differences with respect to phonological structure of the nominal and verbal root. "Nominal lexemes (and consequently denominative verbs) can show a complexity similar to Indo-European," while, in contrast, "[t]he core verb lexicon, depending heavily on a mono-consonantal root, is naturally characterized by frequent homophony" (Rayfield 2002: 1039; Gamkrelidze and Ivanov 1995: 768 also note that the canonical shape of root and affixal morphemes is identical in Kartvelian and Indo-European). Therefore, if the present study investigated the verbal domain, one could expect a rather different behavior of Kartvelian, and such differences in canonical structure between the verbal and the nominal root may well be partly responsible for the weak correlation between the values obtained for the present study and the overall analyzability of lexical items, including verbs, in the comparison with the World Loanword Database data in § 5.2.1.

The typical phonological structure of roots is very different in North-West Caucasian languages. There is little evidence for early contact with other Eurasian language families. This is in contrast to Kartvelian, which shows signs of early Indo-European influence or even co-evolution of lexical items. More importantly, as stated succinctly by Rayfield (2002: 1041), "Abkhaz and Circassian contrast a prodigious wealth of consonants with a paucity of vowels and strict limits on permissible syllable structure. Roots tend to be monosyllabic, sometimes mono-consonantal, consequently with many homophones. Consonants in initial position rarely occur in clusters of more than two, and there are a very limited number of such clusters... As in, say, Chinese, the number of acceptable syllables that can constitute a root morpheme in N.W. Caucasian roots is so small that, in order to express a wide number of concepts or to name, say, flora and fauna, specific lexemes have to be constructed by recombining two or more other lexemes, or otherwise monosyllabic lexemes are polysemantic." The basic facts about Northwest Caucasian phonology and root structure are confirmed by statements of other scholars (among them Hewitt 2008: 307 and Nikolayev and Starostin 1994: 85, 192, who have it that the essentially monosyllabic root structure of Northwest Caucasian languages is due to loss of laryngeals and resonants from the more complex root structures in an earlier North Caucasian stage postulated by them), and is discussed for individual languages of the family. Kuipers (1960: 82-88) provides discussion of the situation in Kabardian, noting in particular the effects the canonical structure of lexical roots has on their semantics. Kuipers (1960: 87) discusses the example of the root *Ŝha* (written later on the same page as *Ŝẖa)*, which ranges semantically over "'head,' 'upper part' (roof, ceiling, summit, seed vessel of flower, ear of corn, riverhead), 'beginning' (of space, of time, crossing of roads), 'important part or member'

(place of honor, head of group), 'spherical part' (bulb), 'covering part' (sleeve), etc., also 'self'." Importantly, Kuipers (1960: 88) also points out that this situation is not much different from the semantic extensions of English *head*, but that still, "the two cases are by no means equivalent, as Kabardian lacks the numerous alternatives with a more limited semantic field that are found in English (roof, top, chief, bulb, etc.), so that polysemy plays a much larger role." Both in Kabardian as well as the sampled Abzakh Adyghe (Paris 1989: 161-162), there are combinations of consonants which act, from the point of view of phonology, as a single phoneme ("groupes consonantiques"). The lexical root in Abzakh Adyghe may consist of a single consonant or a consonantal group as defined above that can but need not be followed by a vowel, or of combinations of the two with insertion of epenthetic shwa (Paris 1989: 163). The apparent pronounced presence of homonymy in Northwest Caucasian languages is at first glance paradoxical, because the number of distinctive consonants is famously high. Thus Abzakh Adyghe only appears to go against the typological trend of having a large consonant inventory and a high degree of analyzability in the nominal lexicon. In fact, phonological restrictions on the level of the lexical root can be held accountable for its behavior. Rayfield's (2002: 1041) further discussion implies that this is less of a problem when it comes to the verbal domain, because the elaborate apparatus of affixation makes it possible to express semantic nuances that are not resolved by the "apparent lexical poverty" of the language, but for the domain of nominals lexical resources appear to be restricted (and note, interestingly, Rayfield's comparison with Chinese!). Concomitantly, Rayfield notes that "[t]he phonological structure of the language and, perhaps, a resistance to alien influences had led, where more sophisticated or abstract vocabulary is concerned, to fewer direct borrowings and more calques" (Hewitt 2005: 139, however, mentions cases of borrowing into Northwest Caucasian languages, but the proportion of borrowings may still be notably lower than in other Caucasian language families). Borrowing behavior is further discussed in § 5.4.2.7.1., but first a brief survey of the situation in the third language family of the Caucasus, Nakh-Daghestanian or Northeast Caucasian, is to follow. Rayfield (2002: 1041) characterizes the structure of the word and lexicon in this language family as assuming an intermediate position between Kartvelian and Northwest Caucasian. He notes, with special reference to the Nakh branch, that the permission of final consonant clusters and the frequency of di- and trisyllabic roots permit a reasonable number of distinct lexical items, while at the same time stating that especially the Chechen lexicon is characterized by a considerable number of homophones. One could thus assume from Rayfield's brief discussion that this intermediate position of the Nakh-Daghestanian family with respect to phonological restrictions on the lexical root inventory would lead Bezhta to have a degree of analyzable terms that is also intermediate between that of Kartvelian and Abzakh Adyghe. However, this is not so; the score for Bezhta is very similar to that of Laz. A quick browse through Comrie and Khalilov (2009a) reveals that most native noun roots in Bezhta have CVC or CVCV shape, which, given the very large inventories of both consonants and vowels,[17] allows for an ample amount of distinct monomorphemic lexical roots. Another reason for

[17] The former is typical of Nakh-Daghestanian languages, while the latter is unusually large compared with other closely related Tsezic languages.

the behavior of Bezhta in the context of the sample that comes to mind is a typological difference between Nakh-Daghestanian and the other two language families of the Caucasus: the former feature noun classes, and the noun class of the arguments are cross-referenced on the verb; this may provide a way to resolve lexical ambiguity on the discourse level (Rayfield's 2002: 1041 discussion also implies this scenario). However, casual inspection of the vocabularies of few or only one language is not sufficiently systematic evidence to show that root structure is a cross-linguistically operative factor. The following section attempts to explore this and related matters more systematically.

5.4.2.6. Canonical Structure of the nominal root

What the situation in the Caucasus shows is that, in general, it would be of great value for lexical typology to have a cross-linguistic study on the possible or typical phonological structure of basic non-derived lexical roots, both for the nominal and verbal domain. While reference grammars of course usually provide information on the syllable canon and phonotactic restrictions, information on the typical structure of lexical roots, and possible restrictions therein is less often found. It seems to be expectable that, if such information were more widely available, they would allow to show strong effects on the structure of the vocabulary, both in terms of the degree of analyzability of lexemes and possibly also on the degree of roots with a comparably vague and broad semantic content. In fact, it is plausible to assume that the interactions between phonology and the degree of analyzability would be further strengthened if this variable could be fully taken into account. As already mentioned, information on the canonical structure of lexical roots is not very frequently provided in reference grammars, but there are exceptions. For instance, Watkins (1974: 74) informs that the canonical shape of nominal (and verbal) roots in Kiowa is monosyllabic and of the shape (C)V(C), where certain consonants can also be followed by the palatal glide /y/, forming a cluster (Watkins 1974: 16), and the final consonant can only be /p,t,m,n,l,y/ (1974: 12-13). Furthermore, Conzemius (1929: 75) states that in Miskito, also a language with a nominal lexicon relatively rich in analyzable terms, "most words have been formed from a comparatively small number of elementary, monosyllabic roots." Miskito, in addition, has a small inventory of distinctive segments, and the basic morphological unit is in fact monosyllabic and the inventory of such units consequently severely limited by phonological factors, so one would expect morphologically complex terms to be relatively frequent in the lexicon, and this is exactly what is observable.

In order to assess canonical root structure in ideally all languages of the statistics sample, also when no such statements are found in the literature on them, the following interim procedure was applied: the number of syllables for all native lexical material in the database not coded as analyzable of any kind or semianalyzable were counted (with anything longer than four syllables, for ease of calculation, being counted as being tetrasyllabic), and then the weighted mean of the count was computed. This provides an empirical measure of the average length of the unanalyzable lexical morpheme in the language in question. However, a problem in obtaining reliable values is that the lexical data are at hand in orthographic, not phonological representation, and the challenge is thus to re-extract phonological structure, in particular syllabification, from orthography. A particularly problematic aspect of this are orthographic sequences of vowels, of which it is

not always clear whether they should be interpreted as diphthongs or sequences of vowels with a syllable break between them. Luckily, frequently such information is available, but there are eight languages, namely Mali, Toaripi, Kildin Saami, Cheyenne, Arabela, Cayapa, Chayahuita, and Cubeo, where vowel sequences in orthography are frequent and their interpretation remains unclear, and another one, Rotokas, where the source (Robinson 2011) briefly discusses the issue of syllabification of adjacent vowels, but remains non-committal as to the correct analysis. Since orthographic vowel sequences of up to five vowel graphemes are quite frequent for instance in Toaripi (Brown 1972: 132), any arbitrary decision as to their treatment engenders the danger of severely distorting the results, and thus, for this particular purpose, the abovementioned languages removed from the sample. In a number of other cases where the interpretation of vowel sequences is an issue, but where they are less pervasive, they were interpreted in a way that disfavors the hypothesis: in languages with a degree of analyzability lower than the cross-linguistic mean, where one would, by hypothesis, expect longer lexical items, they were treated as diphthongs, and in languages with a degree of analyzability higher than the cross-linguistic mean, they were interpreted as sequences. That is, if the procedure is biased in any way, it is biased slightly against the expected outcome.

Another issue is that, of course, the canonical structure of the native lexical morphemes are assessed only on the basis of a very small subset of all nominal items and thus may not be representative. However, for those languages where statements by experts are available, these are in very close agreement with the obtained weighted mean, so that the representativeness of the values seems granted. Table 11 provides these statements, together with the obtained weighted mean.

Language	Expert Statement	Obtained Weighted mean
Mbum	Hagege (1970: 63-64) reports that in his corpus, 55% of lexical items are monosyllabic, and 38% disyllabic.	1.808988764
Ket	"Ket basic vocabulary includes numerous non-derived stems, many of them monosyllabic." (Vajda 2004b: 14)	1.539473684
Carrier	"The primary roots are strictly monosyllabic, and they represent those objects or concepts, which are of the greatest import in American aboriginal life ...the Carrier language could be said to have some affinity to the monosyllabic idioms" (Morice 1932: 24) "In common with the primary roots, secondary roots express concepts or objects of simple import and are likewise unsynthetical substantives; but they are polysyllabic, generally disyllabic, in structure" (Morice 1932: 34).	1.426966292
Kiowa	"there are polysyllabic nouns which can be tentatively regarded as old compounds on the basis of identification of at least one element with synchronically occurring forms. Still other polysyllabic nouns are entirely unanalyzable, but given the monosyllabic structure of roots and the tonal patterns of known compounds, they can safely be inferred to be old compounds" (Watkins 1984: 75-76)	1.3

Itzaj	"Most noun roots are monosyllabic with the shapes CVC, CVVC, and CV'(V)C" (Hofling and Tesucún 2000: 87) "There are also polysyllabic noun roots of the form CVCVC or CVC(V)CVC ... Some of these are undoubtedly derived forms historically but are now considered to be unanalyzed forms." (Hofling and Tesucún 2000: 89)	1.138297872
Hupda	"While Hup strongly favors a syllable-morpheme isomorphism, it also permits words of more than one syllable; these, however, are almost all limited to two syllables. With the exception of ideophones ..., only a handful of words have three or more syllables." (Epps 2008: 80)	1.289855072
Jarawara	"...the language has a strong preference for roots with just two moras ..." (Dixon 2004: 71)	2.525641026
White Hmong	"Le hmong est une langue monosyllabique, les mots, pour l'immense majorité, n'étant formés que d'une syllabe." (Mottin 1978: 4)	1.014925373
Tetun	"Underived lexical morphemes in Tetun have from two to four syllables... most lexical morphemes are disyllabic." (Van Engelenhoven and Williams-van Klinken 2005: 739)	2.114285714

table 11: some expert statements for languages where they are available and computed weighted mean of the canonical structure of the lexical root

There is some areal variation and clusters of each type. For instance, in many languages of Southeast Asia, the canonical root structure is monosyllabic. The map in figure 20 shows the areal distribution of the types.

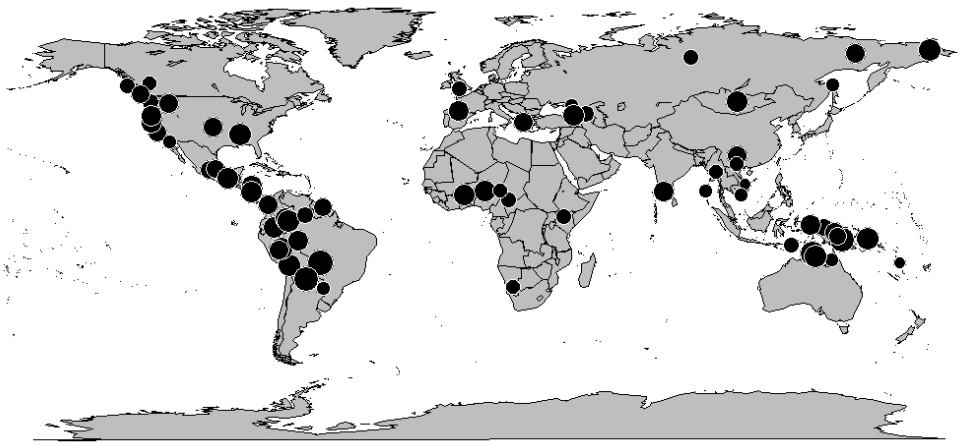


fig. 20: canonical length of the nominal root, reduced statistics sample

The same Mixed Model design employed already for systematic exploration of other phonological features was then used to analyze the data while controlling for area (canonical structure of the lexical stock is, like other phonological features such as those discussed above, susceptible to areal influence; one case in point are the Austronesian languages of the Chamic branch, which have adopted their inherited disyllabic roots to the common Southeast Asian monosyllabic structure, see e.g. Haudricourt 1956). As seen in the plot in figure 21, the same basic tendency already familiar from other phonological features can be observed: lower degrees of analyzability correlate with segmental complexity in nominal roots, and higher degrees of analyzability are found in languages in which the canonical root structure is more simple (the weighted means for each language were partitioned in four groups for visual representation only and are in Appendix C, but the actual more informative values themselves were used for statistical analysis).

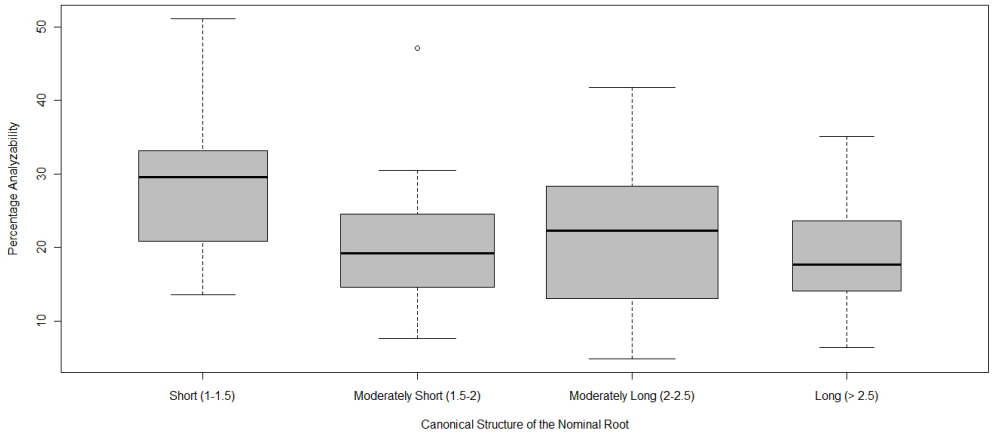


fig. 21: correlation between structure of the canonical nominal root and morphological complexity in the lexicon

Root structure does have a significant impact on the degree of analyzability in the nominal lexicon cross-linguistically when controlling for area (*p*-value for the predictor root structure: .0355, estimate: -5.111). Thus, bearing in mind the difficulties in the assessment of root structure and the hence somewhat error-prone methodology, THE NUMBER OF ANALYZABLE TERMS SEEMS INVERSELY CORRELATED WITH THE LENGTH OF THE CANONICAL ROOT IN MOST REGIONS. Given that this is the last of the features relating to complexity of the sound system and of the word to be discussed, it is now possible to convert the variable as to the type of analyzable lexical item (derived as opposed to lexical) into a cross-classificatory table, the other variable being the number of analyzable items, and add the typological correlate of complexity of the word and of the sound system.[18]

[18] Note that this table simplifies matters in that the degree of analyzabitility and percentage of derived vs. lexical terms are for ease of exposition treated as if these were absolute categories rather than the continua that they actually are.

	High degree of Analyzable Terms	Low Degree of Analyzable Terms
Lexical Dominating, Derived Subsidiary	• Low complexity in verbal person marking, fixed word order •Simple phonology, short roots	• Low complexity in verbal person marking, fixed word order •Complex phonology, long roots
Derived Dominating, Lexical Subsidiary	• High complexity in verbal person marking •Simple phonology, short roots	• High complexity in verbal person marking • Complex phonology, long roots

table 12: updated table showing the correlations obtained so far

5.4.2.7. Two Excursuses

5.4.2.7.1. *Excursus I: The linguistic treatment of items of acculturation, phonology, and overall complexity in the nominal lexicon.* In the discussion of the distribution of analyzability in the nominal lexicon in the Caucasus, it was noted that Abzakh Adyghe features relatively few loanwords when compared with the representatives of the other linguistic families of the Caucasus. Bezhta is rich in loanwords from Arabic, Avar, and more recently, Russian (see Comrie and Khalilov 2009b for full discussion), and Laz features many loans from Turkic, Greek, and Georgian. Further, as seen in § 5.2.2., analyzability in the semantic domains of both nature-related and body-part terms is strongly correlated with that in the domain of artifacts. Thus, one might be lead to hypothesize that the dominant technique a language employs to name novel artifacts, that is, whether it prefers borrowing or coinage of a neologism, is correlated with the degree of analyzable terms present in other areas of the lexicon: languages with many analyzable terms will typically more often accommodate items of acculturation by coining a neologism, while languages with a relatively high degree of simplex lexical items will more often respond by borrowing a name for novel objects from a contact language. Unfortunately, it is not possible to assess this prediction on a global scale on the basis of the sample. This is due to the fact that not all sources indicate the status of the listed lexical items, and it not advisable to attempt to identify loanwords by mere eyeballing, in particular because they are impossible to identify if one is unfamiliar with the donor language(s). Therefore, the discussion is restricted to languages of the Americas and to loanwords of European origin in the domain of artifacts, for two reasons: first, the sources consulted for this area of the world in the vast majority of cases indicate if a given lexical item is in fact borrowed, and should this be not the case, chances are high that loanwords can still be identified as such by inspecting their phonological shape since the donor languages are well-known European languages.

However, a certain margin of error obviously remains, and errors are possible. As elsewhere, it is possible that the same language features more than one term for the same concept, one of which may be borrowed and the other may be native but have experienced semantic extension or may be a morphologically complex neologism. In line with the policy in the overall assessment of morphological complexity, percentages are calcu-

lated, which is the reason why the global values reported in Appendix C are at times smaller than the number of loanwords listed.

Restricting the discussion to the Americas has another reason, namely that a differential degree of borrowing as opposed to coinage has been observed frequently here (e.g. Voegelin and Hymes 1953). The most comprehensive study on the topic is Brown (1999), who investigates the linguistic acculturation in languages of the Americas on the basis of a list for 73 items introduced by the Europeans. For each language in his large sample, Brown studies for how many of these items the languages have borrowed terms as opposed to other strategies of lexical expansion, and provides borrowing scores for each language. Since many languages of the present sample are also represented in that of Brown (1999), a direct comparison is often immediately possible. Relevant data are in Appendix C, where also further information that will become relevant for the present discussion is given: analyzability scores for all meanings except artifacts, as well as information for each language group as to which European power they were in contact with. Figure 22 plots the differential borrowing scores in the Americas obtained by this procedure; these will be used in the following analysis.

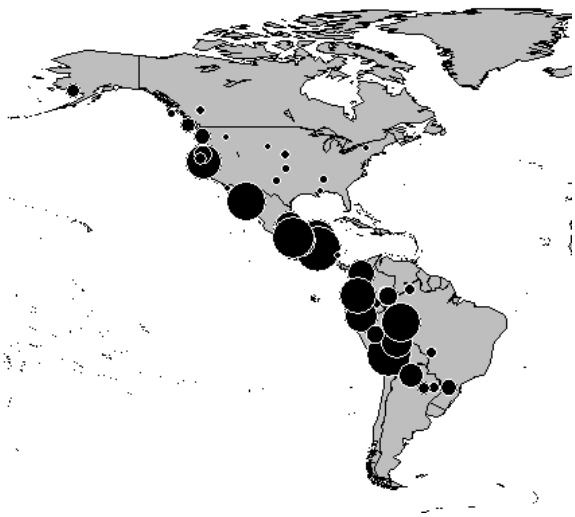


fig. 22: differential borrowing scores in the Americas

There is close agreement between Brown's and the scores obtained here. Although both studies sometimes employ the same source to extract the data, there is variation both in the number of items of acculturation and the individual items they investigate so the present study is not a mere replication of Brown's. What immediately strikes the eye in the map is the differential degree of borrowing depending on what the dominant contact language is. From Brown's data the generalization emerges that languages influenced by Russian, Spanish or Portuguese show a higher degree of borrowing than those influenced by English or French (Brown 1999: 80-81), with languages in contact with Spanish showing

the most pronouncedly high scores. In the words of Brown (1999: 81), "where direct Spanish influence has not been a factor, Amerindian languages have been disinclined to borrow European terms for items of acculturation." Taking up observations made by Bright (1960), Brown (1999: 81-82) relates this fact to the different ways in which Spanish as opposed to English- and French-speaking conquerors treated the indigenous populations which lead to different rates of bilingualism among Native Americans, and which in turn is thought by him to be responsible for the observed differential rates of borrowing. However, the details of the sociolinguistic situations are not elaborated on in depth by Brown.

Brown (1999: 83-91) also devotes space to discussion of the possible influence of structural features of languages on the rate of borrowing. Comparing variation in the degree of borrowing of genetically related languages, it emerges from Brown's study that sometimes languages from the same family, for instance Uto-Aztecan, show marked differences in the degree to which they adopted loanwords for items of acculturation. Like in this case, very often this degree of borrowing is correlated with what the contact language is, in line with the general observations made above: those Uto-Aztecan languages in contact with Spanish-speakers borrowed significantly more heavily than those in contact with English-speakers or French-speakers. For instance, Cora, which came in direct contact with Spanish-speakers, borrowed 80% of terms for the meanings investigated by Brown, whereas Comanche, which has been in direct contact with English- and French speakers and has undergone indirect influence from Spanish only, borrowed terms for only 17% (Brown 1999: 84, table 6.4.). Frequently, where there is little family-internal variation in the percentage of borrowed lexical items in Brown's study, as for Salishan, Siouan, Iroquoian, and Muskogean, it is the case that speakers of these languages had been uniformly exposed to contact with either the English and/or the French, which is further evidence for a scenario in which the dominant contact language is the major factor influencing the degree to which languages integrate loanwords for items of acculturation into their lexicon (by way of hypothesized different rates of bilingualism). Where there are significant differences in the number of loanwords in related languages that have been in contact with the same European languages, Brown tentatively resorts to language purism as an explanation (Brown 1999: 84-85).

A peculiar case is, however, that of the internal variance within the Yuman family. Kiliwa notably receives a loanword score of zero in both the present and in Brown's count, in spite of the contact language being Spanish. Mixco (1977: 20-21) explains the extreme paucity of loanwords to the difficult relations with and the hostility of the Kiliwa to Spanish culture. He also notes that other Yuman languages which are structurally similar to Kiliwa have borrowed more eagerly from Spanish and later also from English, giving figures of "approximately a hundred loanwords" in Diegueño and Paipai and fewer in other Yuman languages. Winter (1992) discusses the situation in Walapai, another Yuman language. Although noting that here there are a few loanwords from English and a somewhat larger number of loanwords from Spanish, Winter (1992: 219) says that "[i]t is widely assumed that Amerindian languages in general make wide use of descriptive terms, that is, of constructs whose parts taken together provide a composite reflection of crucial aspects of the meaning of the term." In Walapai, such morphologically complex terms are rather limited in native vocabulary, occurring most frequently in toponyms. However, in spite of

the comparably limited areas of application of complex terms in native vocabulary, "[i]t was precisely this technique which could be made use of to cope linguistically with a large influx of new notions from the culture of English-speaking Americans, short of taking over a great number of English words" (Winter 1992: 220). Winter's (1992: 222) summary is that the way the language dealt with acculturation was, in spite of a number of loanwords, "a strictly monolingual response in an increasingly bilingual situation."

Brown, in spite of arguing for bilingualism as the primary responsible factor for the differential degrees of loanwords in languages of the Americas, does not entirely rule out the possibility that structural features of languages may influence the degree to which they are eager to integrate loanwords, noting for instance the case of Salishan languages, which have accepted a larger number of loanwords than other North American languages not directly in contact with Spanish (Brown 1999: 90). However, he cautions that integration of lexical items of European origin into Salishan often was indirect via Chinook Jargon, and considers this explanation more plausible than one in terms of structural properties. In summary, Brown's (1999: 91) conclusion is that his data "suggest that if language structure factors affect lexical borrowing, they do so only minimally." That Brown attributes great importance to the contact language and the different sociohistorical circumstances of the contact scenario that come along is convincing, since these factors unmistakably are highly relevant. However, beneath these apparently major factors, there is some variation on a smaller scale that cannot be easily explained and that suggest that something else, even though probably subsidiary, is in play as well.

Though Salishan languages, according to Brown, have a relatively high loanword percentage when compared to other North American languages and this may be due to indirect borrowing via Chinook Jargon, it is still notable that languages spoken on the West Coast, such as Nuuchahnulth and Haida, although incorporating significantly less foreign lexical material than languages that underwent influence from Spanish, tend on average to also score higher on the loanword index than languages of Eastern North America. In § 5.4.2.2., a west-east cline of decreasing phonological complexity and concomitantly increasing analyzability of the lexicon was noted. Could it be the case that languages with a generally analyzable lexicon disfavor borrowing as the prime mechanism of lexical acculturation? This idea has been around at least since Sapir (1921/1970: 195-196), who suggests that resistance to borrowing has something to do with "the psychological attitude of the borrowing language itself." Comparing English and German, Sapir offers a psychologizing account of differences in the structure of the lexicon in terms of the lexicon and hypothesizes effects of these differences on the varying degree of borrowing in the two languages:

> English has long been striving for the completely unified, unanalyzed word, regardless of whether it is monosyllabic or polysyllabic. Such words as *credible, certitude, intangible* are entirely welcome in English because each represents a unitary, well-nuanced idea and because their formal analysis (*cred-ible, cert-itude, in-tang-ible*) is not a necessary act of the unconscious mind (*cred-, cert-,* and *tang-* have no real existence in English comparable to that of *good-* in *goodness*). A word like *intangible,* once it is acclimated, is nearly as simple a psychological entity as any radical monosyllable (say *vague, thin, grasp*). In German, however, polysyllabic words strive to analyze themselves into significant elements. Hence vast

> numbers of French and Latin words, borrowed at the height of certain cultural influences, could not maintain themselves in the language. Latin-German words like *kredibel* 'credible' and French-German words like *reussieren* 'to succeed' offered nothing that the unconscious mind could assimilate to its customary method of feeling and handling words.

Haugen (1956: 66) takes up this idea,[19] and Casagrande (1954: 228) suggests that, next to socio-historical factors, the paucity of loanwords in Comanche is attributable to the fact that "[w]ith an efficient means of word-building at hand, Comanche had little need to resort to linguistic borrowing."[20]

To test the hypothesis of a correlation between a general predilection for analyzability in native vocabulary and the relative degree of loanwords in languages of the Americas, values for the analyzability in the lexicon with the domain of artifacts removed in order to not replicate results were computed (the obtained values can be calculated from appendix B). A Generalized Linear Model, using both the degree of analyzability outside of the artifact domain and whether the dominant contact language is Spanish or Portuguese as opposed to English, French, or Russian was built.[21] Ineseño Chumash, Kashaya, Wappo, Carib, and Miskito, for which English and Spanish influence is about equally strong (though perhaps one of the languages was the dominant contact language at one time, and the other at another time), were removed from the calculation. To rule out possible effects from very closely related languages as well as spatial proximity and therefore potentially highly similar contact situations, only languages from different genera were subject to modeling. Note that this entails that the level of statistical independence is shifted down from the family to the genus level to still allow to include data from as many languages as possible for this particular test. Since the variable presently under investigation cannot be directly influenced by genetic inheritance, this seems appropriate for the present purpose. Data from languages not subject to modeling are presented in italics in appendix C. Modeling was begun by including an interaction factor between contact language and analyzability in native lexicon, which however appeared to be insignificant (p = .5097), suggesting that the parameters are independent or at least do not influence one another when it comes to the respective loanword percentages, and was hence removed from the model. The simpler overall model is highly significant (adjusted R^2 = .2942, $F_{2, 43}$ = 10.38, p = .0002099). As for the individual predictors, there was, unsurpris-

[19] Apparently independently, similar ideas are sketched by Ullmann (1962: 112-113), who also uses German as the example.

[20] Even Mixco (1965: 101) notes that "[n]ominal compounding is a productive syntactic process in Kiliwa" and that this fact "perhaps explains the paucity of Spanish loanwords." Thus, even for the case of Kiliwa, while sociocultural factors are probably the major force explaining the type of lexical acculturation dominant in the language, it may be aided by structural factors of the language.

[21] Note that this test operates with the assumption that loanwords are mostly found in items of acculturation, which in the slice of the vocabulary presently investigated clearly cluster in the domain of artifacts. This, however, does not rule out the possibility that languages also have borrowed from contact languages in other semantic domains, as is the case in some languages of Mesoamerica. Thus, there is the possibility that this fact skews the results in that loanwords, unlike calques, enter the lexicon of the borrowing language as unanalyzable wholes and may have replaced an analyzable native lexical item. A drawback of this approach is that it does not systematically control for this possibility.

ingly, a strong effect of the contact language (estimate: 25.9851, p = .000378), but notably also a weaker effect (estimate: - .9055) of the degree of analyzability that is also significant at p = .020730. Figure 23 plots the percentage of loanwords depending on the degree of analyzability in the remaining semantic domains investigated.

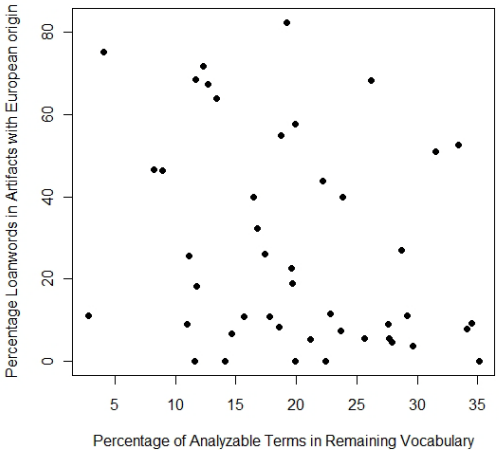


fig. 23: correlation between degree of borrowing and analyzability in the lexicon in languages of the Americas

The conclusion is as follows: IN THE AMERICAS, THE DEGREE OF BORROWING DEPENDS PREDOMINANTLY ON THE CONTACT LANGUAGE, BUT IS ALSO INVERSELY CORRELATED WITH THE DEGREE OF ANALYZABILITY IN THE LEXICON. This is in line with Sapir's statement: languages with an analyzable lexicon less readily accept loanwords than languages that have a larger number of monomorphemic lexical item. To reiterate, this statement should not be read as being equal to denying the overwhelming influence of which contact language is dominant and likely concomitant differences in bilingualism; but below the surface of this obvious difference, there does appear to be a more subtle influence of structural-organizational properties of the lexicon in general that does have an, albeit subordinate, effect on the degree to which a language is likely to accept borrowed terms for items of acculturation.[22] For the time being, the correlations that are obtained can only be said to be valid for the particular case study of the Americas, and it would be necessary to test in greater detail if this situation is demonstrable empirically also in other areas of the world.

[22] Sapir's (1921/1970: 195) position is in fact quite similar: He does not deny that the particular historical circumstances of the contact situation have to play a major role in accounting for differential rates of borrowing, but notes that "it is not the whole truth."

There is some evidence that there is a similar general world-wide trend from the data in the World Loanword Database (Haspelmath and Tadmor 2009c). Bradley Taylor (p.c.) kindly computed the simplicity score (as defined in Haspelmath and Tadmor 2009c) for the languages in the World Loanword database excluding loanwords (that is, those lexical items that are coded as clearly borrowed or probably borrowed). In effect, this score reflects the percentage of analyzable lexical items in native vocabulary (though as noted already above, this score takes into account complex items which are semantically redundant as well as semianalyzable terms). There is a certain trend for languages with relatively low ratios of morphological analyzability to have borrowed more lexical items than those with a higher degree of analyzability in native vocabulary on a global scale (Spearman's $\rho \approx .34$); however this positive correlation fails to reach statistical significance ($p \approx .11$).[23] It is plotted in figure 24.

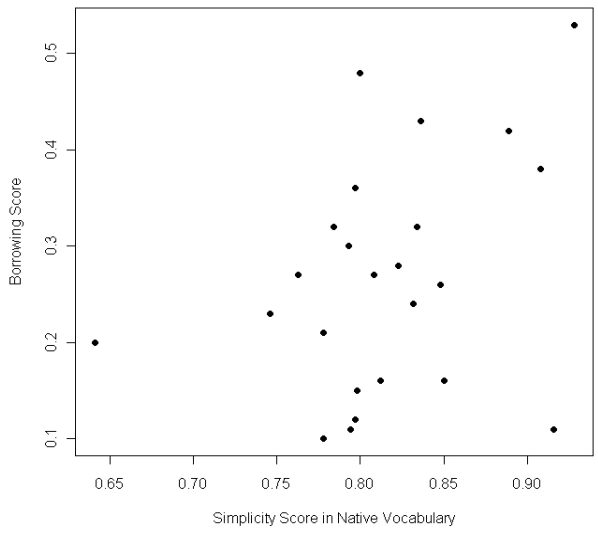


fig. 24: Simplicity Score in Native Vocabulary and Borrowing Score; data from the World Loanword Database (Haspelmath and Tadmor 2009c)

Thus the evidence from evaluation of the data from the World Loanword database is somewhat inconclusive, and it would require more in-depth research to either confirm or refute Sapir's (1921/1970) statement as to the influence of a relative paucity of morpho-

[23] Some of the languages in the World Loanword database are members of the same language family. To avoid possible biases from structural factors and to allow for statistical testing, one language per family was selected at random and the Creole languages Seychelles Creole and Saramaccan were excluded from analysis. The languages which entered calculation are, with their simplicity scores discarding loanwords following in parentheses: Bezhta (.784), Ceq Wong (.908), English (.889), Gurindji (.8), Hupda (Hup) (.797), Imbabura Quechua (.793), Japanese (.797), Carib (Kali'na) (.85), Kanuri (.778), Ket (.778), Kildin Saami (.823), Mandarin Chinese (.641), Mapudungun (.832), Oroqen (.916), Santiago Mexquititlan Otomi (Otomí) (.794), Q'eqchi' (.798), Swahili (.795), Takia (.834), Berber (.928), Thai (.848), Vietnamese (.808), White Hmong (.746), Wichí (.812), and Yaqui (.763).

logically complex terms in the lexicon and the predilection for accepting loanwords on a global scale empirically. In the above discussion, this was done by comparing the degree of analyzable terms in a selection of native vocabulary items. Indeed, one way to assess the productivity of a particular word-formation device suggested by Plag (1999) is to simply measure how many lexical items were created by its application. However, it is just one way, and since accounts such as Casagrande's (1954) explicitly refer to the productivity of the word-formation apparatus, rather than to the degree of analyzability in the conventionalized lexicon, a worthwhile investigation would be to also assess morphological productivity in other ways (see Plag 1999 for a number of suggestions as to how to measure productivity of derivational affixes in English). At any rate, a quite tentatively further correlate related to the degree of analyzable terms can be added, namely the differential rates to which the languages rely on borrowing as opposed to coinage of complex neologisms from the native stock of lexical items.

	High degree of Analyzable Terms	Low Degree of Analyzable Terms
Lexical Dominating, Derived Subsidiary	• Low complexity in verbal person marking, fixed word order	• Low complexity in verbal person marking, fixed word order
	•Simple phonology, short roots	•Complex phonology, long roots
	•Tentatively: favors neologisms	•Tentatively: favors borrowing
Derived Dominating, Lexical Subsidiary	• High complexity in verbal person marking	• High complexity in verbal person marking
	• Simple phonology, short roots	• Complex phonology, long roots
	•Tentatively: favors neologisms	•Tentatively: favors borrowing

table 13: updated table showing the correlations obtained so far

However, it must be emphasized that this result is tentative only and awaits further in-depth investigation.[24]

5.4.2.7.2. *Excursus II: A note on analyzability in Proto-Indo-European and other Eurasian reconstructs.* In Proto-Indo-European, the reconstructed ancestral language of the Indo-European languages, the canonical structure of the lexical root is monosyllabic (Szemerényi 1990: 130). The canonical Indo-European root is of CVC, or better CeC structure, with the consonant qualities being fixed and therefore root-defining and the vowel quality subject to systematic ablaut. The root can be augmented by resonants to yield structures such as *CReC, *CeRC, *CReRC, with **i* and **u* being capable of acting as resonants. In

[24] Rice (2012: 70-71) for instance, rejects internal structural factor as the cause of the high degree of motivation in the Athapascan language Dene Sųłiné, instead, inspired by Thurston (1989), arguing for little bilingualism as the more likely relevant factor (compare also § 5.4.2.12.1).

addition, there are constraints on possible root structure: roots cannot contain two plain voiced stops, or a voiceless stop and a voiced aspirate (Fortson 2004: 72, Szemerényi 1990: 99). As established by Benveniste (1935), the Indo-European root may be further expanded by a consonantal element ("root determinatives," "root extensions," "root enlargements") to form a stem of either verbal or nominal nature, yielding the form CVCC called 'theme I' and, for verbs only, CCVC called 'theme II.' Thus the root **pet-* 'fly' with the suffix *-er* yields **pét-r-*, continued in Sanskrit *pátra-*, and **pt-ér-*, continued in Greek *pterón* (Szemerényi 1990: 131); in both cases the meaning of the stem is 'wing.' Since roots are sometimes augmented by a preceding **s-*, of which it is sometimes unclear what governs its presence or absence (the so called s-mobile), the monosyllabic lexical morpheme of PIE can actually become complex with up to five consonants, with a CCCVCC structure, but "[e]ven apart from these reduced forms obtained by removal of the root determinatives, it can be empirically established that the majority of the monosyllabic roots contain only two consonants with the basic vowel *e* between them" (Szemerényi 1990: 131) and that "the structure of most PIE roots can be boiled down to a single template, **CeC-* (Fortson 2004: 70; CeC is, however, only the canonical root structure, and a number of roots with *a*-vocalism as well as non-canonical shape are found, Szemerényi 1990: 132, Fortson 2004: 72). As Lass (1994) points out with respect to the various extensions which need to be posited by the evidence from the daughter languages as augmenting canonically shaped roots, extensions of the root could be viewed as the "detritus" of old word-formation devices, the precise function of which cannot be recovered, given that a reconstruction of the PIE situation which posits that the extensions are part of the roots and cannot be segmented is unparsimonious in that it posits numerous synonymous and partly homophonous roots. What is more, Iverson and Salmons (1992) suggest, partly on grounds of typological naturalness, that even CVC root structure in Proto-Indo-European reflects a relatively late stage in the development of the language, with the consonant in the coda originally augmenting a simpler CV-type syllable structure and fused with the root already in the stage of the language that posits canonical CVC structure.

Importantly, many of these basic underived lexical roots within the lexicon of Proto-Indo-European are verbal in nature, with the root determinatives serving to derive both nouns and verbs, and further enhancements "always produce noun stems" (Szemerényi 1990: 131).[25] The above examples of roots augmented by a determinative, **pét-r-* and **pt-ér* 'wing' from the root **pet-* 'fly' already provides the transition to this aspect of the PIE lexicon, since in the case of the reconstructs one is dealing with analyzable terms of the derived type, more precisely, derived from a verbal root. In fact, Wodtko et al. (2008: xvi) note that "[g]erade deverbal motivierte Nomina stellen einen umfangreichen Teil des gemeinsamen indogermanischen Wortschatzes dar, da das urindogermanische Lexikon in stärkerem Maße deskriptive Mittel verwendet zu haben scheint, als es in vielen modernen

[25] The situation is in general in marked contrast to the situation in the neighboring languages of the Uralic family, in which disyllabic verbs and nouns or noun-verbs abound, with inflection and derivation obtained by suffixation (Janhunen 2001: 209). Uralic will be dealt with briefly later.

indogermanischen Sprachen der Fall ist"[26] / "deverbally motivated nominals in particular constitute a substantial part of the common Indo-European leixcon, as the Proto-Indo-European lexicon seems to have used descriptive means to a larger extent than is the case in many modern Indo-European languages," and Nichols (2010: 47) therefore calls PIE a verb-based language. Given that deverbal nominalizations are semantically and morphologically dependent on a verb, Wodtko's (2005: 50-51) conclusion is that they play a marginal role in the lexicon as mere makeshift devices that can when required be coined ad hoc, need not be learned, and are easily understood by way of being related to a verbal root. The important question as to the degree of conventionalization of deverbal nominalization which obviously cannot be answered for a reconstructed language put aside, this is a matter of the point of view one takes: if they are indeed frequent, then it could also be said that they, or rather the mechanism of nominalization per se, plays a major role in the organization of the PIE lexicon.

Unanalyzable nouns (that are by virtue of this of course also not deverbal) are, however, clearly also reconstructible for PIE. One type of athematic root nouns includes terms for "core vocabulary" meanings such as *h_3ekw- 'eye,' *ped- 'foot' and *dem- 'house' that probably represent an old stratum of the lexicon (though note that Rix and Kümmel 2001: 297, 458, as cited in Wodtko 2005: 63, posit verbal origins even for the terms for 'eye' and 'foot'). Like athematic nouns, there are also instances of nouns in the other major class of Indo-European nominals, thematic nouns, that are not relatable to other roots, among them generic level terms for animals and kinship terms such as *$u̯lk^wos$ 'wolf,' *$h_2ŕ̥tḱos$ 'bear,' *$snusós$ 'daughter-in-law' and *ag^wnos 'lamb' (Fortson 2004: 116; Wodtko 2005: 70-72 also mentions body-part, kinship and fauna terminology as the semantic domains in which monomorphemic nouns in PIE are found, see also § 5.4.1. for typological comparison). However, the majority of thematic nouns stand in a derivational relationship to known roots (Fortson 2004: 116, see Fortson 2004: 116-118 for an overview of noun-deriving processes). Wodtko et al. (2008: xiv) note that also the PIE root is capable of acting as a free-standing form, but still the root nouns ("Wurzelnomen") can be seen as an abstract or agent nominalization of the corresponding verb, see also Fortson (2004: 108-109) for an overview. Another type of root noun forms agent or undergoer nouns from verbal roots (Fortson 2004: 109); it appears to be these that Wodtko et al. (2008) are talking about.

Indeed, there are many unanalyzable lexical items in modern daughter languages, including many in "basic" vocabulary that can through comparative historical work be traced back and linked to stems based on typical CeC roots. Further, if the reconstructions are accurate, many PIE vocabulary items for the meanings on the wordlist used for the present study were analyzable in PIE, more precisely deverbal derivatives. Some assorted examples include those in table 14.

[26] A footnote by Wodtko et al. (2008) refers to Seiler (1975), whose work and elaboration on the notion of "descriptivity" was discussed in chapter 2. In fact, Seiler (1975: 38-39) briefly comments on the relationship between the frequently transparent relation between arguments and a predication by virtue of many nominals being derived from verbs and thus 'describing' their referent. He also suggests that this structure might be correlated with the absence or optionality of the copula in older Indo-European languages.

Root and gloss (original glosses in square brackets)	Derivative	Cognate of derivative (information in parentheses added)	Reference
*$b^herĝ^h$- 'become high, arise' ['hoch werden, sich erheben']	*$b^herĝ^h$-*o*-	Germanic **berga* (German *berg* 'mountain')	Wodtko et al. (2008: 30-31)
*h_1ed- 'bite, eat' ['beißen → essen']	*h_1d-*ont*-	Germanic **tanþ* (German *zahn* 'tooth')	Wodtko et al. (2008: 208, 210)
**sed*- 'sit down' ['sich setzen']	**ni-sd-ó*-	Old High German (and Modern German) *nest* 'nest'	Wodtko et al. (2008: 590-591)
*$h_2eḱ$- '(be/become/make) sharp, pointed' ['scharf, spitz (sein/werden/machen)']	*$h_2ák$-*mon*-	Lithuanian *akmuō* 'stone' ['Stein']	Wodtko et al. (2008: 287)[27]

table 14: examples of PIE deverbal derivatives

Of course, this is merely impressionistic and anecdotal evidence, and what would actually be required to allow for systematic exploration is a full 160-item wordlist for PIE, but the impression that the table above gives receives backup by experts on Indo-European as underscored by the quote from Wodtko et al. (2008) cited above, although one problem noted by Wodtko (2005: 52) are methodological difficulties in deciding whether a given derivative with reflexes in daughter languages does indeed entail that the derivative must be posited for the Proto-Language, since it could also be possible that the template for word-formation rather than the resulting form may have been inherited and daughter language terms coined independently on the basis of the common template.

Be that as it may, further questions that arise are: how natural is such a lexicon in which analyzability seems to be so pervasive cross-linguistically, and do other aspects of Proto-Indo-European as presently reconstructed accord with this observation to form a harmonic whole? With regard to semantics of the roots and lexical items derived from them, a lexicon as reconstructed by Pokorny (1959/1994), in which highly abstract meanings are dominant, are unnatural and implausible typologically (Sweetser 1990: 25-27), and these apparent shortcomings are likely due to the lack of a principled methodology of semantic reconstruction that does not generate a large number of highly abstract meanings for roots such as 'to swell' or 'to be bright' which abound in Pokorny (1959/1994) for reconstructs (a problem noted by Rix 2002: 1336). Put strongly, one could even say that a lexicon with such reconstructed semantics is a violation of the uniformitarian principle.[28]

But what about the sheer quantity of analyzability, regardless of the naturalness of semantic structure found in analyzable terms? Could a higher degree of analyzability in the Proto-Language also be an artifact of reconstruction, that is, does the very process of historical reconstruction of earlier stages of the lexicon of related languages necessarily

[27] Note also the PIE term for 'stone' mentioned in the very beginning of Chapter 1. This term is not mentioned by Wodtko et al. (2008).

[28] For instance, in the reconstructions proposed by Jóhannesson (1949) for PIE body-part terms, there is a conspicuously large number with a literal meaning of 'the curved one' or 'the swollen one.'

involve the discovery that synchronically unanalyzable lexemes in many cases can be traced back to morphologically complex ones? In a sense, this seems to be trivial, since complex terms are the norm rather than the exception for novel terms (Hagège 1993: 182-183 among many others), but is it a necessary concomitant of reconstruction, given the fact that after all one of the very task of etymological research is to make synchronically unanalyzable terms transparent by putting them in diachronic perspective (Rix 2002: 1336)?

At this point, the typological correlations established so far may help. In § 5.4.2.6., it was suggested that there is a correlation between the canonical structure of the lexical root with the degree of analyzability to the effect that the shorter the canonical root is, the more analyzable terms are found in the languages of the sample. Further phonological evidence is also available: assuming a standard non-glottalic reconstruction of the PIE consonant inventory with about 25 distinctive segments (15 stops in three series -voiceless, voiced, voiced aspirated- and five places of articulation -labial, dental, palatal, velar, labiovelar-, fricative *s, liquids *l and *r, nasals *m and *n, glides *i̯ and *u̯, and three laryngeals *h_1, *h_2, and *h_3, which would be an average-sized consonant inventory in terms of Maddieson 2005a), it becomes clear that the number of distinct roots with canonical shape this inventory is able to generate, not least due to the prevalence of *e*-vocalism, is clearly restricted; probably not as severely as the Vanimo system with the figure of 960 distinct morphemes calculated by Ross (1980), but also not unimaginably large (Jucquois 1966 counts about 2,000 attested roots from Pokorny 1959/1994). When it comes to the meanings expressable by these roots, the same is obviously true, and Jucquois (1966: 65, table 2) shows that the number of homophonous roots is very high, effectively reducing the number of 2,000 attested roots with distinct meanings to a much smaller number of attested roots with different phonological shape.

Thus, relating the evidence as to PIE root structure to the typological correlation between canonical root structure, size of the consonant inventory and analyzability in the lexicon, it is no surprise to find that the PIE lexicon appears to have been characterized by a high degree of analyzable terms. In general, leaving aside questions of details of reconstruction and the naturalness of the heavily root-based morphology of PIE, what the present study furthermore demonstrates is that A NOMINAL LEXICON THAT IS CHARACTERIZED BY ANALYZABILITY TO A DEGREE AS THAT APPARENTLY FOUND IN RECONSTRUCTED PIE IS NOT A TYPOLOGICAL ODDITY, which one might be inclined to think judging from the impression gained when comparing the reconstructed stage of PIE with modern daughter language or other better-known European languages, but has parallels in other languages of the world (see Comrie 1993 for discussion of the role of typological naturalness in historical reconstruction). As far as the aspects presently under investigation, a language like Kiliwa is typologically somewhat similar to Proto-Indo-European: an average-sized consonant inventory, with monosyllabic roots dominating the entire nominal and verbal lexicon, including a number of nonanalyzable nouns with this structure (see also § 5.4.1), but a large amount of nominals with more complex structure being either synchronically derived from verbs by a variety of morphological means or at least diachronically relatable to them (although the

nature of the derivational processes differs to some extent) and a high degree of analyzability in the lexicon in general.[29]

Another question that arises is: how did it come about that many modern daughter languages seem to be characterized by a markedly lower degree of analyzable terms when compared with their reconstructed progenitor? On the one hand, this observation hardly requires a special explanation, since it is an ubiquitous process for erstwhile morphologically complex terms to become phonologically reduced and demorphologized, in short, lexicalized as single unsegmentable wholes. On the other hand, there is evidence from at least two subbranches of Indo-European, Germanic and Slavic, that typological shifts took place in the lexicon which may have supported the transition from a largely analyzable nominal lexicon to a more unanalyzable one. Nichols (2009b) shows that in Slavic a lexical type shift from verb-based to noun-based took place. Kastovsky (2006a, b) demonstrates that in Germanic, there was a shift for the base form on which inflectional and morphological processes operate from the root-based type found in PIE to a stem-based type. This shift came into being by word-level stress becoming fixed in Germanic which made formerly predictable ablaut alternations unpredictable on the one hand, and on the other by an increase of secondary derived nouns and verbs. This ultimately led to the emergence of a new stem unit which served as the input of derivational processes (Kastovsky 2006b: 163). Still later, loss of medial and final unstressed syllables which were morphologically speaking markers of grammatical information occurred, ultimately leading to word-based morphology in Modern English. Of course, a logical concomitant of this development would be that inherited root-based derivatives would be reduced in their transparency in a perhaps more pervasive fashion than in the case of garden-variety lexicalization processes in individual items, as the productive apparatus of word-formation shifts to being based on the stem with derivational morphemes becoming reinterpreted as belonging to the stems.

Let us now turn to the question as to the potential artificiality of Proto-Language analyzability as a by-product of reconstruction, by comparing the reconstructed PIE state with that of two other Proto-languages of major language families of Eurasia, Uralic and Nakh-Daghestanian (in principle, it is not important that the language families are also located somewhere in Eurasia, and other language families might have been adduced as well). The reconstructed phoneme inventory of Proto-Uralic is somewhat smaller than that of PIE, with about 20 consonant phonemes (Rédei 1988: ix). As to the structure of the morpheme, demonstratives are reconstructed as monosyllabic and content words with very few exceptions as disyllabic, with the subtypes VCV, CVCV, VCCV, CVCCV, VCCCV, and CVCCCV (Rédei 1988: xi). Like Indo-European, Uralic is a deep family, with a primary split between the Finno-Ugric and Samoyedic subgroups. Janhunen (2009: 68) tentatively suggests a split of Proto-Uralic at 5,000 BP, but even the Finno-Ugric branch is assigned the proposed age of 4,500 years by Janhunen – plenty of time for diachronic change, in-

[29] One aspect not mentioned so far is that, as discussed for instance in Fortson (2004: 122-123), compounds are also reconstructible for PIE, and thus one could speculate that PIE belonged to languages of the mixed type as defined in chapter four for the present study; this would be another parallel to Kiliwa, in which both analyzable terms of the derived and lexical type are found.

cluding possible lexicalization of erstwhile morphologically complex lexical items to occur. Probing the Uralic lexicon for such processes at random using Rédei (1988) as a resource does indicate some phonological reduction and monosyllabification occurring in daughter languages when compared with the Proto-Uralic or Proto-Finno-Ugric reconstruct (one typical component of lexicalization). Importantly, however, there is no indication that in the reconstructed state of affairs, the parent term was morphologically analyzable, but rather appears to have been a unanalyzable word following the canonical disyllabic Uralic root structure. Table 16 illustrates this point, using the same set of four meanings listed above for Indo-European. In the case that states are reconstructible for several genealogical levels, the one for the highest level was selected as an example.

Meaning	Reconstruction	Cognates (selection, some marked as tentative by Redei); original glosses in square brackets	Reference
'mountain'	**kaδ'a*	Hungarian *hëgy* 'mountain, tip' ['Spitze; Berg'] Tas dialect of Selkup *ķée* 'hill' ['Hügel']	Redéi (1988: 115)
'tooth'	**piŋe* (Proto-Finno-Ugric level)	Finnish *pii* 'tooth, spike, peg, outer corner of house' ['Zahn, Zacke, Stift; äußere Hausecke'], Estonian *pii* 'spike, tooth, prong; sinew, muscle' ['Zacke, Zahn, Zinke; Sehne, Muskel'], Birsk dialect of Cheremis *püj*	Redéi (1988: 382)
'nest'	**pesä*	Finnish *pesä* 'nest,' Kildin and Notozero Saami *piess* 'Vogelnest,' Ezrä-Mordvin *pize*, Hungarian *fészëk* 'bird's nest; seat, abode' ['Vogelnest; Sitz, Wohnsitz]	Redéi (1988: 375)
'stone'	**pije*	Finnish *pii* 'firestone' ['Feuerstein'], Chantaika dialect Jenissej-Samojedic *fû*, Tawgy-Samojedic *fála*, Motor *hilä*	Redéi (1988: 378)

table 15: Some Proto-Uralic reconstructions and cognates in modern languages

While in this sense the report in table 15 is selective, it is not selective in that examples which do not involve erstwhile morphological complexity were deliberately chosen - the same situation is found in terms not listed in table 15, and thus there is no evidence that present-day simplex lexical items can be reduced on a larger scale to analyzable proto-language equivalents.

Nakh-Daghestanian is another ancient Eurasian language family. Nichols (2003: 297) considers the family to be at least 6,000 years old so that the age of this language family is also comparable to that of Indo-European (if not a bit older). The consonant inventory reconstructed for Proto-Nakh-Daghestanian is complex, and nominal root structure was canonically disyllabic, allowing for consonant clusters. Roots were required to contain at least one obstruent. In most languages of the Daghestanian branch initial consonant clusters are not allowed or can be shown to be secondary in some cases where they are found, but the reconstructed situation is preserved in Nakh languages (Nikolayev and

Starostin 1994: 82). Using data from Nikolayev and Starostin (1994)[30] for the same set of four meanings, the situation is in fact parallel to that found in Uralic, as seen in table 16: some phonological reduction in a number of daughter languages, but no evidence for erstwhile morphological complexity on the level of the proto-language.

Meaning	Root and Gloss	Cognates (selection)	Reference
'mountain'	**muɦialV* 'mountain'	Chechen *lam*, Avar *meʕér*, Archi *mul*	Nikolayev and Starostin (1994: 834)
'tooth'	**čɬɦV* 'tooth'[31]	Ingush *carg*, Bezhta *sila*, Tabasaran *slib*	Nikolayev and Starostin (1994: 326)
'nest'	**mōngwē* 'nest; bed'	Karata *minge*, Akushi Dialect of Dargwa *muga*, Lezgian *mug* 'nest, burrow; basket, hive; tree-hollow'	Nikolayev and Starostin (1994: 828)
'stone'	**hrŏmçwe* 'stone'	Botlikh *hinça,* Lak *nuwçi* 'iron or stone plate for roasting grain,' Khinalug *riçin*	Nikolayev and Starostin (1994: 495)

table 16: Some Proto-Nakh-Daghestanian reconstructions and cognates in modern languages

Again, the examples are meant primarily for illustrative purposes, and there are more synonymous or near-synonymous reconstructed lexical items not listed here which, however, also are not or do not appear to be morphologically complex. Notably, in addition, even for the rare cases where trysyllabic forms need to be reconstructed for Nakh-Daghestanian (such as **ʔVmx̱wĕlʔē* 'wild turkey,' Nikolayev and Starostin 1994: 225), there is no statement in the source that these are due to morphological complexity. Thus, there is no indication from the admittedly somewhat casual inspection of the reconstructed state of affairs that historical reconstruction necessarily leads to the establishment of Indo-European-style word families connected by a shared (verbal) root, and in this case, through this fact the reconstruction of PIE morphology, word structure and deverbal derivation gains plausibility precisely because it is not some inherent property of the method that is the cause for the reconstructed state of affairs.

If indeed PIE was a language characterized by a high degree of analyzability in the nominal lexicon as the evidence suggests, then this finding can be taken as an incentive to speculate about the behavior of the language in other related areas and thus to bring to

[30] Note that Nikolayev and Starostin (1994) entertain the controversial hypothesis that at an even deeper time depth Nakh-Daghestanian and Northwest Caucasian languages are genetically related ("North Caucasian"). For the present purpose, the search was restricted to lexical items reconstructible for the level of Proto-Nakh-Daghestanian to avoid making any commitment as to the accuracy of the claim of genetic relatedness between the two language groups.

[31] This root is reconstructed for Nikolayev and Starostin's "North Caucasian" level; it is offered here since no separate reconstruction for the Proto-Nakh-Daghestanian level is provided by them, although this should surely be possible.

light other aspects of the linguistic prehistory of Indo-European. One deverbal nominalization in PIE mentioned by Wodtko (2005: 61) is the word for 'plough,' *h_2arh_3-*tro*-*m*, consisting of the verbal root *h_2arh_3- 'to plough,' the instrument nominalizer *-*tro* and the nominative case suffix *-*m*. Wodtko further notes that a noun for this artifact that is independent of the verb was apparently not available to speakers of PIE, and adds in a footnote that this demand was also not met by borrowing from a contact language at the time of the Proto-Language. While this certainly is at first glance a trivial contingent fact about Indo-European, it is possible to actually ask the question: why not, and can this behavior be motivated? If indeed borrowing behavior should turn out to be related to analyzability in the native lexicon on a global scale, then one could expect the common ancestor of Indo-European languages to have had a dispreference for borrowing, but rather to have preferred coining neologisms (probably a considerable number of them by derivation from verbal roots) for artifacts using native lexical material. Thus, in this case, unfortunately for the task of establishing the areal context in which PIE was spoken by inferencing from loanwords, one could tentatively suggest that one should not expect to find much evidence for language contact as evidenced by apparent loanwords in PIE, since it would be natural for the language, given its typological characteristics, to have favored descriptive neologisms over loanwords. Leaving aside the vexing issue of the identification of the direction of borrowing, which can at times be even hard to determine in the case of actually spoken language, and even more so in the case of a reconstructed prehistoric language, there is some evidence for borrowed lexical material in the PIE lexicon. Gamkrelidze and Ivanov (1995: 769-776) even make a commitment as to the direction of borrowing, by stating that there are a number of loanwords from Semitic and Sumerian in PIE that predominantly denote domesticated animals and cultivated plants as well as names for particular tools and numerals, alongside loans from PIE in Kartvelian and languages of the Ancient Near East (some of the purported Semitic loanwords in PIE, such as the word for 'star,' are controversial however). If indeed there are loanwords in PIE, then this does not devalidate the hypothesis, which merely states that the number of loans should be relatively small. Note, however, that there are many ifs in the above statements; for one thing, the proposed account operates with the assumption that the situation found in the Americas is indeed replicable on a global scale which has not been demonstrated presently, and further, as already stated above, it would require a more systematic exploration of a larger portion of PIE vocabulary to consolidate the very fact that it was characterized by a large number of analyzable terms for a number of standardized meanings. Still, the case of PIE shows how typological data based on synchronic observations has the potential to contribute to questions of historical linguistics that are set quite removed from the present date, even though it can never be the only piece of evidence to solve puzzles of linguistic prehistory, which always requires detailed work by philologists.

5.4.2.8. Interactions between individual predictors

This section takes up the main thread of this chapter after the excursuses. Having established four apparently relevant phonological or morphophonological factors, namely size of the consonant inventory, complexity of the syllable, tonality, and length of nominal roots, it is important to assess whether these are independent of one another cross-

linguistically or linked in some way. In discussing the effect of differences in syllable structure complexity and consonant inventory size in § 5.4.2.2., Maddieson's (2005d) suggestion as to an interdependency between the two variables was pointed out. While in the words of Maddieson (2005h: 15), "absolutely no correlation was found between the number of vowels and the number of consonants" (see also Justeson and Stephens 1984 for a full-length study), Maddieson (2005d) notes that there is a correlation in his sample of 484 languages between the structure of the syllable and the consonant inventory size, to the effect that with increasing complexity in the syllable structure there is a rise in the mean of consonant inventory size in the languages of his sample, as seen in table 17.

Syllable Structure	Average Number of Consonants
Simple	19.1
Moderately Complex	22.0
Complex	25.8

table 17: Average number of consonants for languages with different levels of complexity in syllable structure, adapted from Maddieson (2005d)

However, as acknowledged by Maddieson (2005d: 55) himself, his sample is neither controlled for genetic nor areal effects, and thus he cautions that the results may be due to fortitous historical contingencies rather than a genuine "design feature of language." Furthermore, in the discussion of the diachrony of tone, a correlation between tonality and complexity in the syllable structure as well as the structure of the lexical root were alluded to.

Given these different suggestions as to interactions between the variables that play a role in shaping the structure of the lexicon, it is imperative to test in a systematic fashion what correlations exist between the relevant variables concerning complexity in phonology and root structure in the present sample to obtain a better understanding as to which phonological features are really relevant in shaping the degree of analyzability in the nominal lexicon. When assessing dependencies of consonant inventories and syllable structure putatively identified by Maddieson (2005d) on the basis of the languages in the statistics sample, which has the property of being genetically balanced, indeed such a dependency is found (Spearman's $\rho = .3$, $p = .01276$; in this case, unfortunately no model also taking into account areal factors is possible because the residuals do not fulfil the required assumptions). The associated plot is in figure 25.

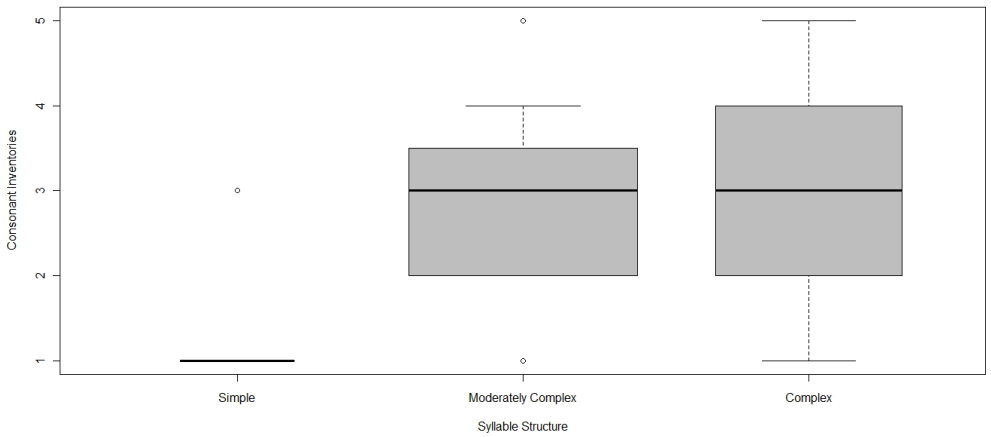


fig. 25: Size of consonant inventory correlated with complexity in syllable structure

Thus, at least speaking for the sampled languages, the two measures are not independent (and Maddieson's larger sample suggests that this might be also true on a larger scale). Now, this fact does not damage the findings regarding the effect of phonological factors on the lexicon: rather than acting as independent factors influencing the structure of the lexicon, one could then say that they "team up" and together exert influence on the degree of analyzable lexemes in the lexicon. But which factor, if any, is more important?

Moreover, as visualized in figure 26, canonical structure of the nominal root was found to be predictable (p = .022, estimate: -.1061) by the size of the consonant inventory: the larger the consonant inventory is, the shorter are the lexical roots. This correlation is similar to that found by Nettle (1995, 1998) based on smaller samples.

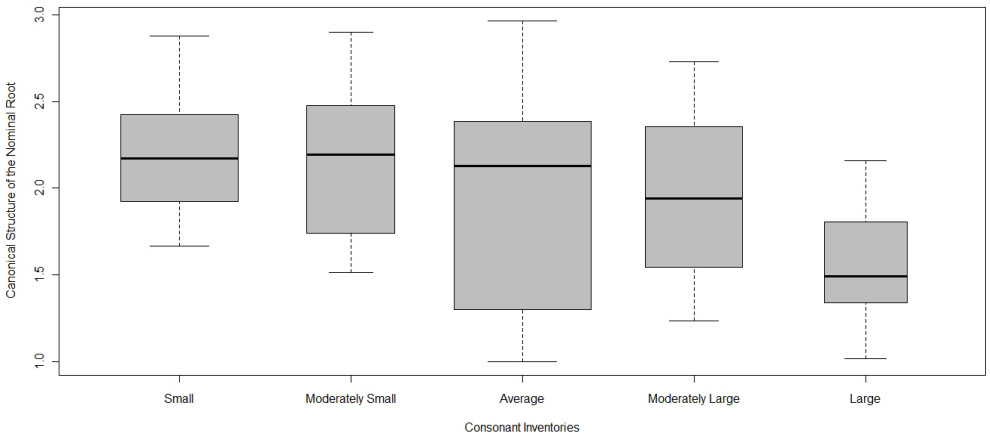


fig. 26: Size of consonant inventory correlated with canonical structure of the nominal root

Moreover, unsurprisingly given the suggestions in the literature, canonical root structure is predictable by tonality (p = .0014, estimates: -.3615 and -.6817). In languages with complex tone systems, length of nominal roots drops dramatically (not just in Southeast Asia), as seen in figure 27.

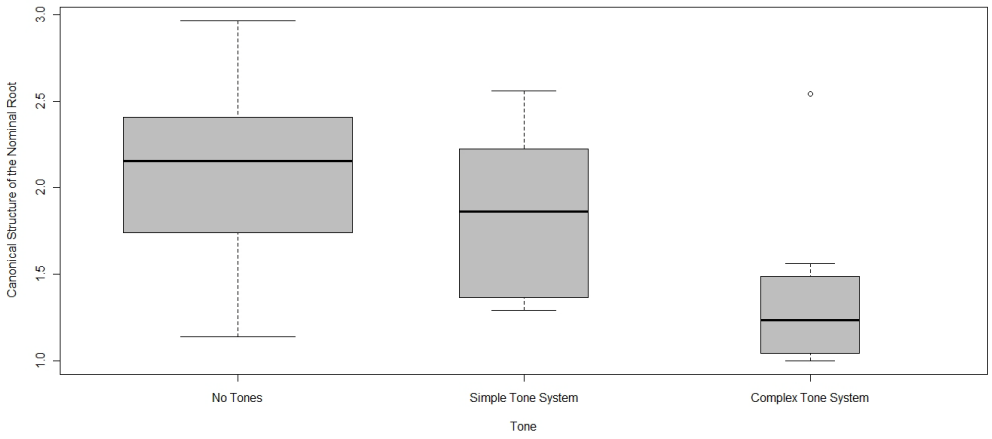


fig. 27: Canonical structure of the nominal root correlated with tonal complexity

Thus, summing up the evidence so far, there are four to some extent interrelated phonological factors interacting with the degree of analyzability in the nominal lexicon to a statistically significant degree: the size of the consonant inventory, which is itself correlated positively with complexity in syllable structure and negatively with the canonical shape of the nominal root. This latter factor in turn interacts with tone, to the effect that when tonal complexity increases, roots become shorter. Figure 28 summarizes the dependencies diagrammatically, with black arrows between features indicating a dependency.

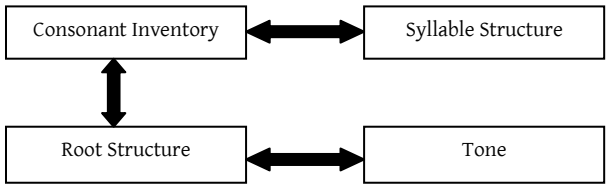


fig. 28: correlations between different aspects of morphophonological structure

Obvious and important questions that arise from these findings are thus (i) whether it can be assessed if one or a subset of the phonological properties is really the relevant one for the behavior of the sample languages with respect to analyzability, with the other(s) being a side effect due to interactions in phonology that are independent of this and (ii) how

precisely the features interact with each other, i.e. whether there are combinations of feature values that give rise to particularly high (or low) degrees of lexical analyzability.

To answer these questions, however, more complex statistical analyses would be called for, for instance a General Linearized Mixed Effects Model. While this is not in principle a problem, it is an issue because the sample size of the present study is relatively small, and thus not all logically possible combinations of values are attested in the sample. This already becomes a problem when trying to take into account only two features. For instance, there are no languages in the sample with complex tone systems and simple or moderately complex syllable structure, there are no languages in the sample with large consonant inventories and a simple syllable structure, and so on (and for many, but not all, other combinations of values there is just a single language in which it is realized). When combining all three relevant features, the coverage becomes even more fragmentary, and there are very many combinations of values which are simply not attested in the sample. This is a situation that is detrimental for the reliability of statistical analyses and the conclusions that can be drawn from them, because statistical power of the model is then low and it becomes instable in that very small changes in the data can have dramatic effects.

What the dependencies between the individual variables however at any rate do show is that there is every reason to believe that the features interact in significant ways, and that their effects combine in exerting influence on the structure of the lexicon. In the absence of reliable possibilities of statistical testing, this can be shown in the following fashion: when values for the segmental phonological variables that showed significant interaction in the lexicon are combined to a single index (bypassing tone, both because the decrease in analyzability as tone systems become complex is hard to interpret and because here the correlation is positive and thus hard to integrate into a combined measure with the otherwise consistently negative correlations), effects become very strong. Combining the individual variables is done by conflating the information they provide into one variable, which will be called the COMBINED PHONOLOGICAL COMPLEXITY INDEX (CPCI) in the following for want of a better term. The CPCI is computed in the following way: first, the value for the canonical shape of the native unanalyzable lexical morpheme is scaled down to ordinal scale with four levels of variation (as is done for the plot in figure 21). Languages with values between 1 and 1.5 are grouped together, and so are those with values between 1.5 and 2, 2 and 2.5, and 2.5 to 3. This entails that some information is lost, but the procedure is statistically valid nevertheless. Now, there are three variables: consonant inventories with five levels, syllable structure with three levels, and root structure with four levels. In order to normalize the different scales and thus to render the values comparable, they are multiplied to reach the smallest common denominator, which is 60. Thus, the value for consonant inventories is multiplied times twelve, that for syllable structure times twenty, and that for root structure times fifteen. These values are then added up and the sum is divided by the number of attested values, ideally three, but sometimes only two due to lack of secure data (if only one feature value is available, the CPCI is not calculated). Values for the CPCI are in Appendix C. In a Mixed Effect Models, with the percentage of analyzable terms as a response value, the CPCI as a fixed effect and area as a random effect, there is a very significant impact of the CPCI on the analyzability score at $p < .0001$. Figure 29 plots the results.

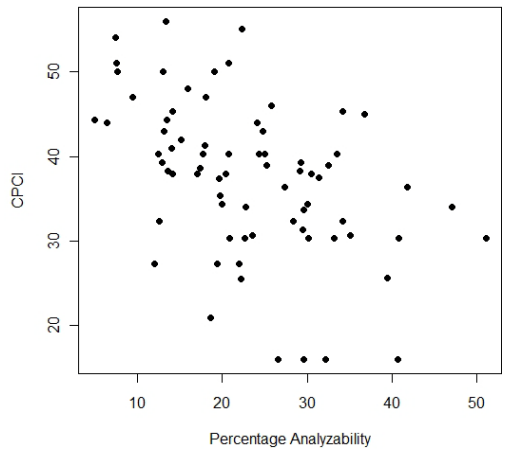


fig. 29: correlation between the combined phonological complexity index and analyzability

Thus, a combination of the individual features leads to a much stronger correlation with analyzability in the lexicon than when each is taken as a fixed effect on its own. And this is hardly surprising, given that languages such as Pawnee and Bororo, which have both simple syllable structure, small consonant inventories and relatively short lexical roots, are situated on the upper end of the continuum in the number of languages with analyzable terms.

5.4.2.9 Intra-family variation in phonological complexity and analyzability

It is also possible to ask whether the same principle that is operative typologically is also observable within language families (see Greenberg 1978, 1995 for intra- and intergenetic comparison respectively). Given that genealogically related languages started out from a common state with respect to the phonological system, complexification or simplification in that system would be expected to have an effect on analyzability in the lexicon. And if this prediction turns out to be true, it would be evidence for the operation of some sort of diachronic pressure that causes languages to adapt with respect to one of the variables as the other changes.

For this purpose, CPCIs were computed for all languages in the entire (EXT-2) sample which (i) fulfil the criterion of having more than 65% of equivalent terms for the meanings available for comparison and (ii) are known to go back to a common ancestral language. Criterion (ii) was applied rather strictly in this context, that is, the comparison was only carried out if the genetic relationship is firmly established and not controversial among experts of the families in question (for this reason, for instance, Kanuri was not compared with Dongolese Nubian and Ngambay since the genetic link as parts of Nilo-Saharan between them, in particular concerning Kanuri, is not uncontroversial; similarly, comparison was not carried out between languages classified as Pama-Nyungan in Dryer

2005a, although the general trend to be reported there is also observable here, and neither were "Hokan" languages compared). The obtained results are in Appendix C. What they show is that, for the eleven language families for which only two languages are compared, in seven, the language with the higher CPCI has the lower number in analyzable items, and that with the lower value in the phonological complexity index the higher degree of them. Data from three families, Afro-Asiatic, Jivaroan and Quechuan, run counter to this trend, while the evaluation for Nilo-Saharan is equivocal since both relevant languages receive the same score for the CPCI. Where more than two languages enter into the comparison, values are not always paired hierarchically, but here, correlation tests are available to assess the dependency between the two variables. This turns out to be always negative, as expected. It is relatively weak in the case of Niger-Congo (Spearman's ρ = -.08571429), and quite strong in Austronesian (Spearman's ρ = -.6107894), Sino-Tibetan (Spearman's ρ = -.5), and Uto-Aztecan (Spearman's ρ = -.5). Thus, THE SAME COVARIATION BETWEEN PHONOLOGICAL AND LEXICAL COMPLEXITY THAT IS OBSERVABLE IN AN INTER-FAMILY TYPOLOGICAL COMPARISON IS ALSO OFTEN NOTICEABLE WITHIN LANGUAGE FAMILIES THEMSELVES.

However, there is the issue that sometimes, the languages of the same family are quite heterogeneous typologically, and in addition, their sociolinguistic status may be vastly different (take, for instance, Manange vs. Mandarin; for the same reason, arguing on the basis of the Austronesian data is dispreferred due to differences in societal scale). While the hypothesis presently entertained is that phonological factors alone are dominantly responsible for the variation in analyzability, there may be other yet undetected interfering factors in the play (see also § 5.4.2.12), and these may cause the correlation to be altered in either direction. One may think of diverging grammatical organization, but also differences in sociolinguistic setting, including factors such as the number of speakers, the size of the territory they occupy, whether or not the language is learned as a second language, etc. Therefore, a particularly useful test case would be one in which both differences in grammatical as well as social structure are reducible to a minimum so that other factors are unlikely to play a big role, the only major difference between the languages of the same family lying in the phonological system. In other words, the variation should be confined exclusively to the variables in question, with everything else being as similar as possible (such an approach in comparative linguistics is first applied by Pederson 1993 and taken up by Bickel 2003). The languages in the sample that come closest to this ideal situation are Cayapa and Tsafiki, both members of the small (five languages) Barbacoan language family of lowland Ecuador and Columbia. The language family is sketched in Curnow and Liddicoat (1998), and the following information is distilled from their account unless otherwise indicated. Cayapa has 3,000 speakers, roughly 20% of them being bilingual, and Tsafiki 1,000 speakers (2,000 according to Dickinson 2002: 20) with about half of them being bilingual in Spanish. Grammatically, all Barbacoan languages, including Cayapa and Tsafiki, have a similar profile: they have SOV word order, are predominantly suffixing, and have alignment systems on a nominative-accusative basis. In fact, Cayapa and Tsafiki are very closely related even within Barbacoan, jointly forming the South Barbacoan subgroup of the family, and yet, they differ in the size of their consonant inventory. Cayapa has an inventory of twenty-four distinctive consonants (*p*, *t*, *tʲ*, *k*, *b*, *d*, *dʲ*, *g*, *ts*, *tʃ*, *f*, s, *ʃ*, *h*/*x*, *s*, *m*, *n*, *ŋ*, *ɲ*, *r*, *l*, *ʎ*, *w*, *j*, and *ʔ*, with the phoneme /g/ being marginal and proba-

bly introduced into the language with Spanish loans). In contrast, Tsafiki only has fifteen consonants (*p*, *t*, *k*, *b*, *d*, *ts*, *Φ*, *s*, *h*/*x*, *m*, *n*, *r*, *l*, *w*, *j*, and perhaps *ʔ*, cf. Moore 1972). In both cases it must be noted that the languages are phonologically quite underanalyzed and the systems in their synchronic state are therefore somewhat insecure. The reconstructed Proto-Barbacoan phoneme inventory, from Curnow and Liddicoat (1998: 401, table 9), is in table 18.

	Consonants				Vowels	
p	t		k	i	ɨ	u
	ts					o
Φ	s	ʃ	h		a	
m	n					
	l					
	r					
w	j					

table 18: Proto-Barbacoan phoneme inventory (Curnow and Liddicoat 1998: 401)

Thus, rather than Tsafiki having shrunk its consonant inventory when compared with Proto-Barbacoan, it is rather the case that the Cayapa inventory expanded by phonemicizing erstwhile allophonic differences which are still observable synchronically in Tsafiki. Moore (1962) reconstruct both palatal and alveolar series for Proto-South Barbacoan, i.e. the common ancestor of Cayapa and Tsafiki, which Curnow and Liddicoat (1998: 400) show to be unnecessary. Allophonic variation was phonemicized when Cayapa collapsed *o and *u, leading to the emergence of a new series of palatal consonants *ʃ*, *tʃ*, *ɲ*, *ʎ*, *tʲ*, *dʲ* next to the alveolar series which all Barbacoan languages feature (Curnow and Liddicoat 1998 leave the development of voiced stops, *s* in Tsafiki, and both *r* and *s* as well as that of the palatal stops in Cayapa unaccounted for, but this does not alter the synchronic observation that these contrast exist in the present-day languages, no matter how they arose). Moreover, *ʃ becomes *s* in Cayapa and is also lost in Tsafiki.

Not discussed explicitly by Curnow and Liddicoat (1998) is syllable structure. In Tsafiki, there are only CV syllables (Moore 1972: 76), and vowel sequences are separated by epenthetical glottal stop, i.e. there are no diphthongs (Dickinson 2002: 34). In Cayapa, CVC syllables are allowed, but the final consonant can only be a nasal, liquid, or glottal stop (Moore 1962: 273 also reconstructs this state of affairs for the common ancestor of Cayapa and Tsafiki). Reconstructions by Curnow and Liddicoat (1998: 392, table 1) suggest that Proto-Barbacoan allowed for CVC syllables, with little apparent restrictions on which consonant can be present in coda position. The coda restrictions in Cayapa are explained by the loss of word-final stops in both Cayapa and Tsafiki.

Summing up, syllable and root structure were simplified somewhat in Cayapa, while Tsafiki has shifted entirely from moderately complex CVC syllables to permitting simple CV syllables maximally. Concomitantly, the Cayapa consonant inventory expanded, and that of Tsafiki underwent some changes when compared to the Proto-Barbacoan state, but remained largely constant in terms of sheer size. Both facts converge in the same predictions about the lexicon in line with the typological evidence: expansion of the

Cayapa consonant inventory may have caused the number of its analyzable terms to shrink, while Tsafiki should have expanded the degree of analyzable terms in its lexicon due to the shift from CVC to CV structure of the syllable/root. The monosyllabic structure of many lexical roots inherited from the Proto-Language remained intact in both languages (albeit they are of different complexity, see above).

Looking in more detail at the individual analyzable terms in Cayapa and Tsafiki, there are a number of meanings expressed by analyzable terms in both languages. Some of these have the same internal semantic structure, so that, other things being equal, they should probably be taken to be inherited at least from the immediate common ancestor. These include 'guts' (Cayapa *pe-shilli* 'exrement-line,' Tsafiki *pe-silí* 'excrement liana/rope') and 'nostril' (Cayapa *quij'juru ~ quijuu ~ quij'jura* /quijcapa-juru/ 'nose-hole,' Tsafiki *quinfu foró* 'nose hole'). Further terms that are quite similar in their internal structure are those for 'ashes' (Cayapa *ñiipe* /ñi-pe/ 'fire-excrement,' Tsafiki *nin fu* 'fire feather/body.hair') and 'brain' (Cayapa *mishpe* /mishu-pe/ 'head-excrement,' Tsafiki *fu-pe* 'hair-excrement'). In the case of other meanings, both languages have analyzable terms, but with different structure.

There are also meanings which are expressed in Cayapa by an analyzable term, but not in Tsafiki. For instance, Cayapa has *ya-tape* 'house-grass' for 'nest' and Tsafiki the unanalyzable *ta'sén ~ ta'sín*. However, it is much more frequently the case that it is Tsafiki which features an analyzable term, whereas the Cayapa counterpart is either totally unanalyzable or semianalyzable. For instance, Cayapa has *ujtupe* 'dust,' Tsafiki *to poyó* 'earth/soil smoke/cloud/steam,' Cayapa has *pusu* 'lake' (< Span. *pozo*?), Tsafiki *hua pipilú* containing *hua* 'big' and *pi* 'water, liquid, river,' Cayapa has *ingbi* 'saliva,' Tsafiki *pi'pí*, presumably reducplicated from *pi* 'water, liquid, river,' etc.

Given that the relevant languages started out as being the same language (or dialect continuum) with the same or highly similar phonological and lexical structure, this is evidence that there is some structural pressure working in diachrony that causes the lexicon to adapt to subsequent phonological developments at some point of time after breakup of the proto-languages, and this appears to be the case not just in Barbacoan, but given similar results in other families, also elsewhere. What is the nature of this pressure? By asking this question, the discussion enters into the last phase of the progress towards an explanation in terms of Bybee (1988): first, *empirical generalizations* were made concerning interdependencies between four morphophonological factors, then, not the least by the computation of the CPCIs, a *principle* was formulated that summarizes several empirical generalizations, and now, the principle needs to be accounted for and an *explanation* for its operation must be sought for.

5.4.2.10. Towards a functional explanation

5.4.2.10.1 *Narrow explanation in terms of homonymy avoidance*. Linguistic universals, and for that matter, presumably also universal tendencies and correlations, are ultimately diachronically motivated and the outcome of some sort of structural or cognitive pressure pushing languages to behave in certain ways, but not in others (Greenberg 1978, 1995, Bybee 1988, Payne 1990, Haspelmath 1999, Bickel 2007, 2008). As Bybee (1988: 351) says, "synchronic states must be understood in terms of the set of factors that create them."

What the case study of Polynesian, the discussion of the situation in Proto-Indo-European, and especially the case study of Mandarin Chinese have shown is that small phoneme inventories or inventories in the process of shrinking may cause problems due to the reduced expressive possibilities, and, in drastic cases, a high number of homonyms in the lexicon when viewed in synchrony, or, when conceived of from a diachronic point of view, the creation of homonyms from erstwhile distinctive lexical items.

In fact, there is a principle said to work against this, namely homonymy avoidance: "[a]ny change in which homophony (words with different meaning sounding the same) is avoided or eliminated" (Campbell and Mixco 2007: 20). Homonymy avoidance is invoked in both synchronic phonological studies to motivate the presence of certain phonological rules, as well as in diachrony to explain aberrant phonological or lexical change, the general assumption being that ambiguity of this kind causes disturbance in the one-to-one match of form-meaning relations impeding successful communication, and that linguistic systems are designed in ways to avoid such disturbances (Plank 1981: 165, who also notes that a generalized theory of ambiguity with predictive power is lacking). The following section provides an overview of research on this, and discusses whether or not the principle of homonymy avoidance is a viable and convincing (diachronic) functional explanation for the observed correlations.

Synchronic phonological studies recurring to homonymy avoidance include the following: Awóbùlúyì (1992) demonstrates that some dialects of Yoruba, including the standard variety, have innovated a rule by which monosyllabic low-toned verbs are required to show mid tone before polysyllabic object-NPs when they are also specified for number and person. Hence, in dialects not having this rule, *mo fò díẹ̀* 'I jumped a little' and *mo fò díẹ̀* 'I skipped some' are homophonous, whereas the standard variety and relevant dialects have *mo fò díẹ̀* 'I jumped a little' versus *mo fo díẹ̀* 'I skipped some.' In Comaltepec Chinantec, a languages in which most words are monosyllabic and in which tone therefore has a high functional load in the lexicon, tone sandhi is rampant, and yet sandhi processes are almost always allophonic and do not neutralize contrastive values required to maintain distinctiveness of lexical items (Silverman 1997). Similarly, in Korean, where neutralization of contrasts is pervasive, these create a very small amount of homophony, and other plausible and phonologically natural neutralizations that would have such an effect are not part of the phonological system (Silverman 2010). In both cases, Silverman explicitly argues that the phonology is sensitive to contrast maintenance. Accounts in terms of avoidance of homophony frequently pertain to grammatical paradigms rather than the lexicon per se. In the Trigrad dialect of Bulgarian, vowel lowering in unstressed syllables is blocked if grammatical endings are present which would produce homophony (Crosswhite 1999). According to Lyovin (1977), in Classical Tibetan, gaps in verb paradigms occur when the future form would be homophonous with the present form, and that such clashes are avoided by the use of periphrastic constructions. In Carrier, diachronic vowel syncope was inhibited in a valency prefix which would have caused it to become homophonous with another one (Gessner and Hansson 2004), and in Banoni (Austronesian), erstwhile distinctiveness of vowel length was gradually lost, except to maintain distinctiveness of bare nouns and their possessed (1st person) counterparts (Blevins and Wedel 2009: 152-154).

The amount of diachronic studies in which homonymy avoidance plays a role is even more numerous. In spite of the Neogrammarian claim by Osthoff and Brugman (1878: 107) that "[m]assenhaft Beispiele beweisen ... dass die Sprache niemals aus Scheu vor Formenzusammenfall oder um Formendifferenzierung zu erhalten Lautgesetze in ihrer Wirkung inhibiert" / "copious examples prove ... that language never inhibits sound laws in their operation for fear of collapse of form or to maintain differentiation of forms," there is a wealth of literature attempting to demonstrate that just this is the case, leading Campbell (1996: 77) to state that avoidance of homonymy as a functional principle in diachrony "is an undeniable empirical reality." The case for such a principle was first made (or at least first popularized) in an oft-quoted study by Gilliéron and Rocques (1912), see Williams (1944: 23-44) for discussion of still earlier precursors. These authors famously observed that, while reflexes of Latin *gallus* 'rooster' are found throughout Southern France, they are notably unattested in Gascony. Here, 'rooster' is denoted by terms that originally meant 'pheasant' or 'vicar' and the inherited word is lost. Now, the area where this lexical replacement has taken place coincides very well with a sound change merging word-final [l] with [t]. Due to this change, Latin *gallus* 'rooster' would not be reflected as *gal*, as in most other areas, but as **gat*, which also happens to be the regular reflex of Latin *cattus* 'cat,' and their argumentation is that this replacement is motivated by the avoidance of homonymy between 'rooster' and 'cat,' two meanings expressed both by nouns and likely to co-occur in the same (rural) setting, thus endangering the successful transmission of information. Similarly, Öhmann (1934: 40) attributes replacement of *fliegen* 'to fly' by *fahren* 'to ride, go' in some varieties of German to clash with *fliehen* 'to flee.' Williams (1944) discusses, next to a wealth of other cases, the fate of English *ear* 'ear' vs. *near* ~ *ear* 'kidney.' Simplifying Williams's more complex discussion, in Northern England and Scotland *lug*, a word of unclear provenience, came to be used for 'ear,' while Standard English *ear* swamped out *nere* ~ *near*, not in the least due to additional confusion when *ear* is preceded by the indefinite article. Discussion of Dutch examples is in Kieft (1938); such early discussions are heavily inspired by Gillierón and Rocques (1912), and indeed their account has spawned much literature that attempts to unravel similar cases in other languages. Dworkin (1993a, b) shows that in Old Spanish, one of two competing same- or similar-sounding lexical items in the same syntactic category and with similar or opposed meanings were lost, and Malkiel (1952) discusses cases on the basis of data from Spanish and other Romance languages. In Proto-Aztec, reflexes of Proto-Uto-Aztecan *$t\dot{\imath}^{n}$ 'stone' and *$t\check{a}$ 'fire' would have been expected to fall together in **te* due to merger of the vowels *$\dot{\imath}$ and *ă. However, the actually attested reflexes are *tle-* 'fire' and *te-* 'stone' in some dialects and *ti-* 'fire' and *te-* 'stone' in others (namely those lacking *tl*), and this case of irregular sound change is explained by Campbell (1975), from who the discussion is summarized, by appealing to the principle of homonymy avoidance. Campbell and Ringen (1981) and Campbell (1988) provide an overview of further cases from the literature where homonymy avoidance is claimed to cause lexical loss or replacement, such as loss of Middle English *quean* 'low woman' after merger of of middle english [ɛ:] and [e:] due to conflict with *queen* except for dialects of the Southwest where the vowels did not merge (taken from Menner 1936: 232-233), replacement of *fliege* 'fly' by *mücke* 'gnat' in dialects of German because of homophony with *flöhe* 'fleas' (taken from Bach 1969: 168). Exceptions in sound

changes in grammatical paradigms are also at times attributed to homonymy avoidance, cases in point being the non-systematic retention of intervocalic *s in Classical Greek due to its function as a marker of the aorist (Bloomfield 1933/1984: 362-363) and that of word-final *n as a marker of the 1st person singular in northern Estonian (Raun and Saareste 1965: 62); for summarizing discussion of both see also Blevins and Wedel (2009) and Campbell (1975).[32]

Malkiel (1979: 2-3; 7) lists four possible outcomes of homonymic clashes (see also Williams 1944 for a similar typology): (i) both lexical items may simply continue to coexist, (ii) one ousts the other (as argued for by Gilliéron and Rocques 1912 and in subsequent studies), (iii) if a semantic gap can be perceived, they may merge (traditionally known as contamination, discussed also in Malkiel 1952), and finally (iv) they may differentiate in form (and meaning). Akin to the last point is the cause of irregular sound change or the blocking of regular sound change highlighted by Campbell (1975, 1996, 1998) and refined by Blevins and Wedel (2009). Furthermore, there is a fifth possible strategy, alluded to by Rédei (1970: 11): therapeutic borrowing, which involves borrowing of a word for one of the referents expressed by homonyms from a contact language. Haspelmath (2009: 50) also mentions that it has been suggested that the replacement of English *bread* 'roast meat' (< Old English *bræde*) by a loanword from French (namely *roast*, which, incidentally, seems to be ultimately of Germanic origin itself, as evidenced by cognates such as Middle Low German *rosten, rosteren* 'to roast on grate,' Kluge 2002) is motivated by homonymy with *bread* 'morsel, bread' (< Old English *bread*), though he remains agnostic as to whether this is really the functional motivation, referring to Weinreich (1953: 58) who uttered a similar opinion. However, there is a sixth apparent possible outcome, pointed to already by Öhmann (1934), and this is more relevant in the present context: creation of a disambiguating compound. Öhmann discusses the case of Middle High German *mûl* 'snout' (an inherited word) and *mûl* 'mule' (< Lat. *mulus*), which latter survived only in compounds like *mûltier* 'mule-animal,' *mûl-esel* 'mule-donkey' and *mûl-ros* 'mule-horse.' Furthermore, he points out that erstwhile *gift* 'gift' only survives in the complex term *mitgift* 'dower' while the simplex has been ousted, presumably due to conflict with *gift* 'poison.' Similarly, Williams (1944: 11-12) says: "[a] word threatened in its existence by some of the vicissitudes of language development, as, for example, homonymic conflict, may be strengthened, made unambiguous by a modifying phrase or term that is in time considered almost an integral part of the word." Note, however, that in the complex terms discussed by Öhmann, the original monomorphemic homonym is not ousted from the language. Coates (1968) presents a further Germanic case study highly relevant for the present context, inspired by discussion in Kieft (1938). There were three segmentally similar but distinct lexical items in Proto-Germanic: **þĭhstila* 'thistle,' **þinhslā* 'pole, beam, tongue' and **þehsalōn* 'adze.' These remained distinct in older stages of Germanic languages where reflexes are attested (Coates 1968: 470, table 2), but later, putative mutual influence and attrition of the near-homonyms lead to unexpected phonological changes in some daughter languages as well as the irregular collapse of two of the forms for instance in Frisian, where both **þĭhstila*

[32] Croft (2000: 66-68) discusses homonymy avoidance as a possible factor in the evolution in grammatical paradigms, but denies strong effect claiming that tolerance of homonymy is relatively high.

and **þinslā* are reflected as *tiksel*. Summarizing the general outcome of the near-homonymy of the words for the three referents from Coates (1968), there are four major strategies, all but one corresponding roughly to the ones mentioned in the literature. In four Germanic languages, there are (optional) compounds based on the inherited word of the redundant type. For instance, Dutch has *disselboom*, with *dissel* the reflex of the inherited **þinslā* 'pole, beam, tongue' and *boom* meaning 'tree.' Five languages have resorted to borrowing of a cognate term that is however phonologically distinct to avoid homonymic clash, for instance Swedish has borrowed *dexel* from German. In six languages, semantic change has taken place, either by replacement of inherited terms with more general meanings or by metaphorical extensions of other words. However, importantly, in some languages, terms for one of the meanings have been given up entirely, and replaced by compounds that do not involve one of the inherited words as constituents. For instance, Dutch has *dwarsbijl* 'cross axe,' Yiddish and Faroese have *bonders hak* and *bøkjaraøks* 'cooper's axe' for 'adze,' and Icelandic has *vagnstöng* 'wagon-pole' for the second of the conflicting meanings; Norwegian and Danish have analogous compounds.

Coates (1968) is important in the present context for another reason: he argues that not only perfect homonymy may be a factor, but that near-homonymy is sufficient in some cases to trigger linguistic changes such as lexical replacement, and if this is true on a larger scale, then lexical replacement due to similar forms becomes a more attractive functional explanation to account for the high numbers of complex terms in languages with morphophonologically simple systems, because, while the number of true homonyms may still be limited, the number of phonologically similar lexical items can be expected to be exponentially larger.

Still, there are serious problems in the cogency of applying the complex of data revolving around homonymy or near-homonymy directly as an explanation for the observed correlations. The first question is how disastrous the effect of a particular sound change (in particular phonemic mergers) can be for lexical distinctiveness. This would require detailed investigation of the functional load of a particular phonemic contrast within the lexicon, and to show where this distinctiveness is encroached on by the loss of the contrast. It is intuitively clear that one particular phonemic merger will not affect the lexicon as a whole, but only a well-defined subset. However, as the diachronic studies cited above suggested, already one sound change can lead to changes in the lexicon if it affects a sensitive point therein, namely lexical items that are useful for successful communication (which is after all the job of language) to be kept distinct. Then, clearly, as Lyovin (1977: 121) says in general, the more homophony is produced by sound changes, and the more dramatic they are, the more likely it is by sheer probability that they cause erstwhile distinct lexical items somewhere in the lexicon to collapse even in the same syntactic and semantic class.

One of the two more severe reasons for some skepticism, however, is that, while there are studies that show empirically that semantically redundant complex terms are introduced for the purpose of disambiguation, there is, with the exception of Coates's study, little evidence in the literature that inherited lexical items are given up because of homonymy and replaced by entirely new complex lexemes that do not contain the inherited homonym as one constituent. A further brief but notable comment is that by Shi

(2002: 76) to the effect that in the process of disyllabification of the Mandarin lexicon discussed in § 5.4.2.3.2. "[m]onosyllabic words are replaced by newly created disyllabic words, in other words, earlier monosyllabic words are abandoned," such as *yue* 'concise' by *jian-yao*, and *wu* 'understand' by *li-jie*. Further empirical demonstration of the pervasiness of such developments would be a prerequisite to make a cogent case for avoidance of homonymy or near-homonymy to account for the observed patterns. To be sure, absence of discussion in the literature does not entail absence of the phenomenon and a case made *ex nihilo* cannot be very strong, but if this were a very frequent process, one would assume that it would have been commented on by historical linguists.

What is more, there are also irregular changes that run counter to the putative principle of homonymy avoidance. Dixon (2004: 71) mentions irregular nonce changes from Proto-Arawá to Jarawara, a language which according to him, and in principle in accord with the hypothesis of phonological constraints on the shape of the lexicon, already has a high number of homonyms due to the phonological structure (11 consonant phonemes, four vowels, (C)V syllable structure, yielding 47 possible distinct syllables and thus, given the preference for bimoraic lexical roots, 2,209 possible disyllabic items as calculated by Dixon). For instance, Proto-Arawá had the distinct nouns **ino-ni/ino-ne* 'tooth,' **ini-ni/ini-ne* 'branch,' and **oni-ni/oni-ne* 'name' (suffixes distinguish masculine and feminine forms respectively). By regular change, **ino-ni/ino-ne* 'tooth' became *ini/ino* in Jarawara. However, the reflex of 'branch' is not the expected **ini-ni/ini-ne*, but *ini/ino* as well. In other words, the feminine forms of 'tooth' and 'branch' collapsed phonologically, and according to Dixon, the masculine form of 'tooth' was analogized causing lexical distinctiveness to cease entirely. Furthermore, by the normal diachronic changes **oni-ni/oni-ne* 'name' first became *oni/oni*, but has then undergone irregular metathesis of the masculine form, giving Jarawara *ino*, yielding homonymy of the masculine form with both 'tooth' and 'branch,' and subsequent extension led the feminine form *ini* to extended to cover 'name' (Coates 1968: 473 also observes that "in some cases a minimal distinction is not felt to be worth preserving, that it is regarded as no better than no distinction at all"). In fact, Dixon (1999: 297) even states that "[o]ne characteristic of Arawá languages is a profusion of lexical homonymy, in which speakers appear to delight," and if this is indeed the case, this delight is of course detrimental to the hypothesis that homonymy avoidance is a cross-linguistic valid motivating factor in language change. It cannot be entirely excluded that the wealth of studies on homonymy avoidance as motivating linguistic change are science-historically a result of the seminal study by Gilliéron and Rocques (1912) that sensitivized linguists to the issue and to look for similar cases in other languages (note also that several later authors, e.g. King 1967, called into question the pervasiveness of homonymy avoidance in diachronic change and the existence of therapeutic language change as put forward by Prague circle linguists, while often acknowledging that some changes may be due to homonymy avoidance or more generally are therapeutic measures).

The other great difficulty is that, as Hanks (2000: 206) has it, and as several of the above cited studies (e.g. Williams 1994, Dworkin 1993a,b) emphasize, sheer identity in form between two lexical items does not necessarily constitute a problem, since if they belong to different parts of speech and are semantically remote from each other, they are unlikely to constitute a danger of confusion in actual discourse, so that normal adult

speakers are unlikely to propel therapeutic measures unless perhaps the above requirements are fulfilled, which should be relatively infrequently the case.

A related issue is the personification of "language" as a deliberately acting agent inherent in some accounts (see also King 1967: 850 for critique). It is important not to forget that this is only a metaphor, and that actual speakers, not languages themselves, are the instigators of language change. In this context, a question one must ask is that if homonymy or near-homonymy is avoided cross-linguistically, by whom is it avoided? If it does not seem attractive that native adult speakers should be responsible for linguistic change caused by homonymy or near-homonymy for the above mentioned reasons, it is worthwhile to look at other groups of speakers. Trudgill (2002, 2004), for instance, argues that a considerable amount of homonymy in the lexicon is indeed tolerable for the native speaker, but is unequally more problematic for the language learner (this is part of the argument developed by Trudgill to account for the shrinking of phonemic contrast in Polynesian), since "[t]he less there is to remember, the easier language acquisition is" (2002: 714), which is also taken by him to be the reason of reduced vocabulary size in Pidgins (cf. §§ 5.4.2.12.1. and 5.4.2.12.7).[33] However, when testing for the presence or absence of second language learners by a Mixed Model design (data are in Appendix C), no statistical effect on the difference of analyzable lexical items can be observed on a global scale. Perhaps with more fine-grained systematic data which distinguishes more subtypes and detailed scenarios paying more attention to the sociolinguistics of the language contact situation etc. significant patterns would emerge, but for the time being, there is no evidence for the sheer presence or absence of second language learners on the degree of analyzability.

In contrast, there is evidence in the recent literature on language acquisition that children in learning their L1 have surprising difficulties with homonyms. To be sure, understanding the concept of homonymy requires a lot of cognitive infrastructure. Most importantly, the child has to be able to understand that a referent and the word denoting it are not the same thing and are associated to each other only by convention, and that the conventions are sometimes such that one word may have two (or more) different referents. The relevant infrastructure is developed by age four (Doherty 2004) and yet children have surprising difficulties in experimental settings with homonyms (Mazzocco 1997, Doherty 2004, see further references to earlier literature therein). The difficulties with homonyms may last as long as until the childrens' 10th birthday, that is, until first language acquisition is nearly complete. Striking examples illustrating this are provided in Campbell and Bowe Macdonald (1983). For instance a girl at age 4;3 is shown a number of pine cones and is asked "What are these things?" by an interviewer. The child volunteers the correct answer "Cones." However, when further asked "Where d'you get cones?," the girl answers "At the shop," and when asked, specifies "At Daddy's shop." This answer is surprising, but becomes at least understandable when one knows that her father is the owner of an ice cream shop. Thus, even though the girl clearly knew of the two different referents of *cones*, as evidenced by her volunteering the answer as to the name of the pine

[33] Furthermore, Trudgill (2002) argues that adult bilingualism and learning leads to phonemic simplification and child bilingualism to phonemic borrowing.

cones, she still somehow failed to keep pine cones and ice cream cones apart.[34] Now, children are also extremely creative at making up new words to fill lexical gaps when they have not yet learned the name of an object (e.g. Clark 1981, 1982, 2000, Clark and Hecht 1982). Thus, classical examples of blocking of the application of word-formation rules by an already existent word, such as the lack of an agent noun **better* 'someone who bets' by *better* and **letter* 'landlord' by *letter* (taken from Jespersen 1942: 231) may be suspended for children, in the spirit of Paul (1880/1966: 251): "Die Individuen, welche das Neue zu dem Alten gleichbedeutenden hinzuschaffen, nehmen in dem Augenblicke, wo sie dieses tun, auf das letztere keine Rücksicht, indem es ihnen entweder unbekannt ist, oder wenigstens in dem betreffenden Augenblicke nicht ins Bewusstsein tritt" / "the individuals adding the new to the synonymous old are not considerate of the latter in the moment they do so, it either being unknown to them or at least not entering their conscious mind in the moment in question." For instance, Panagl (1976) reports a child acquiring German deriving the verb *pfeilen* from *Pfeil* 'arrow,' and from *pfeilen*, in turn, the instrument noun *Pfeiler* for 'bow.' Now, if children have problems with homonymy, then it would be logical to hypothesize that it is them who suppress one of the meanings of a homonym by replacing it with a complex novel term which they coin frequently and productively in language acquisition anyway, and if these are taken over by the parents and become institutionalized, then children could be thought to be the propagators of novel descriptive terms, and in the end be identifiable as the agents propelling the correlations in the lexicon. However, to be sure, this scenario is highly speculative, and operates with the unproven assumptions that (i) children in actual life rather than in an artificial experimental setting really have problems with acquiring homonymy as well - here the same observation as for adults may well hold, namely that homonymy may not be a problem for children as well as long as homonyms do not co-occur in the same context and that the words are thus acquired in different conceptual frames with the child possibly not even realizing homonymy, and (ii) that parents propagate children's innovation through the speech community and thus procure conventionalization of the putative innovations.

Since there is no cogent evidence for child or adult language learners as agents in linguistic change with respect to the topic discussed here, it is appropriate to return to adult native speakers for a moment. It is not always the case that speakers can be sure that context will resolve ambiguities in their messages. Charles-Luce (1993, 1997) demonstrates that phonological processes in speech production are sensitive to semantics and pragmatics of the context, in particular that phonemic contrasts in lexical items are preserved more faithfully in semantic and pragmatic contexts where the speaker cannot expect the listener to expect the word to occur in discourse. Shields and Balota (1991) report that the duration of a target word in a sentence was shortest when the target word had already occurred in the same sentence before, somewhat longer when a semantically related word had occurred in the sentence, and longest when the target word was not related semanti-

[34] Note that in a linguist's analysis, the case of *cone* might be treated as a case of polysemy created by metaphorical extensions from '(pine) cones' to '(ice cream) cones' due to similarity in shape. However, for ordinary language users it may be the case that they would not perceive any semantic link between the two referents, in other words, that for them it may be a case of plain homonymy rather than polysemy.

cally to another one having occurred earlier in the sentence. Likewise, Fowler (1988) shows that words repeated in the same stretch of discourse are, compared to unprimed occurrences, shortened in their pronunciation duration by the participants of her experiments, but not when the words are read from a list, and, very importantly for the present context, neither when the words are preceded in discourse by homophonous items, in which case their pronunciation duration is in fact somewhat longer (Fowler 1988: 313)! What this shows is that speakers are aware of potential ambiguities in communicative contexts and actively (though perhaps subconsciously) take countermeasures to make sure to be properly understood, and it may be precisely this fact that is in the end responsible for diachronic effects of homonymy or near-homonymy: irregular sound change due to overly careful, exaggerated pronunciation of relevant lexical items, or their replacement in ambiguous contexts by semantic proxies potentially leading to semantic shift as conventionalization sets in, or their replacement by a circumlocution, which is the most relevant aspect for present purposes. This would be a step to solve the first problem noted above, namely that homonymy is only pernicious if the relevant items belong to the same part of speech and the same semantic domain, as well as the problematic likening of language to a deliberately acting agent.

This explanation is compatible with the old proposal that speakers are caught in between to opposite drives: on the one hand to avoid unnecessary articulatory effort in order to not waste energy, but at the same time have to make sure to be properly understood (Gabelentz 1901, Martinet 1952, Haspelmath 1999, among others). And if these findings are replicable cross-linguistically, then there is a way to escape Haspelmath's (1999) teleological fallacy to take "functional statements as sufficient explanations," by tying the functional statement up with speaker behavior: languages do not have many analyzable terms in order to counter reduced phonological resources, but because speakers introduce them to ensure successfulness of communicative events. However, the main obstacle for a more detailed fleshing out of the precise workings of the principle is at this point of time that, as noted by Geeraerts (2002b: 37) "actual research into homonymy at the level of *parole* is scarce." For this reason, an in-depth discussion which zooms in from the typological bird's eye view to exemplary studies of the actual processes that might operate in discourse to bring about homonymy avoidance is unfortunately scarcely possible.

5.4.2.10.2. *Broad interpretation as a functional continuum.* Apart from homonymy avoidance per se as the functional drive, it is perhaps worthwhile to conceive of actual homonymy, that is, total formal identity, as only the tip of the iceberg of a larger, but less specific pressure exerted by phonological and morphological factors. The evidence presented here suggests that limited phonological resources cause languages to exploit word-formation devices to build their vocabulary to a greater extent. This, in particular when keeping in mind the obtained correlation between the size of the consonant inventory and root structure, is entirely in line with Nettle's (1995) finding of a correlation between size of phoneme inventory in general and mean word length on the basis of a sample of ten languages: the more phonemes the shorter the words, the less phonemes the longer the words. But Nettle does not take into account whether the words have internal morphological structure, so analyzability in the sample lexical items is likely to contribute to Nettle's

findings to some degree, next to the correlation between canonical structure of the lexical root and consonant inventory size mentioned above. In fact, Nettle (1998: 244) argues that "lexical expansion" as a mechanism of adaptation is responsible for longer words in languages with smaller phoneme inventories, which draws near or is even identical to the coinage of morphologically complex words, which are of course, next to being morphologically complex, also longer.

Maddieson (1984: 8) devotes some discussion to possible effects of simplicity in phonological structure for contrastive possibilities, taking the position that there is little evidence of such effects which would include either "unacceptably high incidence of homophony or unmanageably long morphemes." By inspecting dictionaries of languages with very small consonant inventories, amongst them Rotokas and Hawaiian, he concludes that no such consequences on the morphemic level are discernible. An earlier version of the Hawaiian dictionary that is used also for the present purposes, according to Maddieson (1984: 8), states in the preface that the average number of phonemes per morpheme is just 3.5, which Maddieson finds "clearly not unacceptably long." However, note that Maddieson's discussion is concerned with the level of the morpheme, not that of the lexical item, and the properties of the lexicon suggest that here, to some extent morphologically complex items are used for purposes of disambiguation.[35] This is also supported by some amount of semantically redundant complex lexical items in Hawaiian. For instance, *ake* means 'liver' as well as 'to desire, wish, be eager, yearn.' The meaning 'liver' can be singled out by using the compound *ake-paʻa* 'liver-firm' which is "more specific than *ake*" according to lexicographers (and note that 'lungs' in Hawaiian are either called *ake-māmā* 'liver-light,' *ake-makani* 'liver-wind,' or *ake-pāhola* 'liver-spread'). For present purposes, they do not affect the outcome, since complex terms such as *ake-paʻa* are treated as being redundant and are not taken into account, and effects may be more dramatic if they were.[36] A general tendency one would expect on the basis of the correlations established here is that such formally redundant terms are more frequent in languages with low phonological complexity, essentially serving the same purpose as non-redundant complex lexical items, namely to increase lexical distinctiveness when necessary. This would require further testing.

It is also instructive to look at the ratio of potential words that can be generated by the phonological system and the ones actually instantiated. According to Krupa (1966), in Maori, which has a slightly larger phoneme inventory than Hawaiian and Samoan, the

[35] Trudgill (1996: 15) interprets Maddieson to the effect that "it is not the case that languages with small inventories necessarily have longer words, or vice versa," but note again that Maddieson is talking about the morpheme, not the word.

[36] While it is theoretically conceivable to simply add segments to words arbitrarily to enhance distinctiveness, it seems unlikely that an actual speech community, faced with a limited number of acceptable word shapes due to phonological restrictions and canonical roots shapes, agrees by convention to add a sequence of meaningless phonemes to pre-existing words just to increase their distinctiveness. To do so, a much more natural tool is available, namely that to employ the language's word-formation mechanisms to form compounds on pre-existing roots (which then yields a large amount of semantically redundant compounds, as in Mandarin Chinese), or to replace parts of the stock of inherited words by morphologically complex neologisms.

theoretical number of (C)V syllables is 55. 38 of them are attested, representing 67 morphemes which all express grammatical meaning. From these data, Krupa, inspired by similar indices in Greenberg (1960), derives a so-called index of homonymy of 1.76 by dividing the number of morphemes with distinct meanings through the number of syllable shapes. The theoretical number of bi-vocalic morphemes, which presumably bear mostly lexical meaning given the disyllabicity of Austronesian lexical morphemes (Blust 2007), is 3,025. Of these, 1,258, that is, 41.59%, are actually observed. The index of homonymy calculated by Krupa on the basis of a chance sample of 100 items is 2.27, that is, each lexical morpheme in Maori has on average more than two distinct meanings. What this shows is that it is not necessarily the case that in languages with simple phonological systems all potential word shapes are lexically exploited in spite of many lexical items being homonyms (although there are languages where the ratio of attested to possible word shapes is higher, for instance White Hmong, according to Ratliff 1992). This is on the one hand hardly surprising, since after all, speakers do not engage in mathematical calculations of the number of possible words in their languages, and certainly they do not search for phoneme combinations not yet exploited lexically to immediately do so by the mysterious process of *Urschöpfung* at the next best opportunity, given that, after all, the lexical inventory is an organic whole that is for the most part inherited, not created from scratch. But this does not entail that, rather than searching for gaps, which seems unrealistic, speakers resort to the exploitation of word-formation devices for disambiguation of existing homonyms.

Suggesting pressure on the lexicon arising from phonological simplicity is not equal to postulate any principle of grammatical, lexical, or cognitive organization that generally averses homonymous items from the lexicon, homophonous morphemes from grammatical paradigms, or the development of such items in diachrony, in particular if these are limited to relatively few isolated instances (in this following Blevins and Wedel 2009). Rather, what the evidence suggests is that if phonological possibilities are restricted, and the ratio of instantiated lexical items approaches a certain percentage of all possible lexical items generatable by the morphophonological system (not necessarily even close to 100%, as Krupa's 1966 calculation shows), then there is functional pressure on the linguistic systems to develop strategies to counter the limited expressive possibilities constrained by segmental restrictions, either by the introduction of phonemic tonal contrasts (Matisoff 1973), and/or a notable and statistically verifiable increase of morphologically complex items. As Trudgill (2004: 315-316) says with respect to (unguided) second language acquisition specifically, "[t]he problem lies in the relative lack of distinctiveness between one vocabulary item and another, due to the necessarily high level of usage of all possible syllables," but it seems that this is precisely also the tendency that is observable from a cross-linguistic point of view, irrespective of whether the language is learned by L2 speakers or not.

Alternatively, rather than searching for an explicit functional explanation which is operative, the correlation may simply best be viewed as a constant equilibrium, where languages level off at some point on the complexity on the scale, the endpoints of which are extreme simplicity in the lexicon which goes hand in hand with complexity in phonology and the structure of the root on the one hand and dominant analyzability in the lexi-

con accompanied by phonological simplicity and simplicity in root structure.[37] For instance, many languages of Australia populate the niche of the continuum in which complex lexical items are few, but words correspondingly long. Preponderance of an analytic lexicon could be viewed as an attractor, a term adopted by Blust (2007) from Kelso (1995) to account for the remarkable stable disyllabicity of Austronesian languages, and the concomitant countermeasures invoked by individual languages to restore disyllabicity when it is in danger by sound changes (see also Nettle 1995 for an account of the phonology-lexicon interface similar to a self-organizing system).[38] Likely, there is another counteracting tendency to keep memory load within limits (cf. also Fortescue's suggestions discussed in § 5.4.2.12.4) pushing in the other direction: as Lindblom (1998, 2000) argues on the basis of neurological evidence, there are likely memory constraints which favor re-use of already lexically exploited articulatory movements. And then, these are the poles on the continuum of two opposed drives between which speakers of languages are suspended, and which in the end causes languages to level off at some point of the continuum, with none of them cross-linguistically favored, but jointly defining the space of cross-linguistic variation. Thus, this account would be a slightly modern version of the old notion of poles of articulatory ease and communicative efficiency between which speakers are suspended (Gabelentz 1901, Martinet 1952, Haspelmath 1999), and which cause the languages they speak to be spotted on some place of the continuum.

A similar scenario to the one developed here is also outlined by Nettle (1999: 144), who specifically also mentions the arise of homophony as a result of loss of segments in large phonological inventories due to difficulties to distinguish adjacent segments with similar, but not identical, articulatory and acoustic properties in actual discourse:

> as a result, sets of words that were previously distinct become homophones. When words have become homophones, speakers may have to compensate by some kind of lexical strategy, such as coining a new word or paraphrase. ... Discrimination failure leads to smaller inventories, and the lexical strategies by which meaning is maintained tend to produce longer word forms. The pressure on the language from discrimination failure thus precisely balances that due to articulatory economy. The actual system of any given language emerges from a dynamic equilibrium between these two factors.

However, as noted above, actual evidence that new words are coined to avoid homonymy in particular, is not totally lacking, but relatively sparse.

However that may be, such scenarios are interesting also in light of the recent surge in interest in linguistic complexity, beginning with McWhorter (2001) and challeng-

[37] A phonological simplicity/complexity continuum, or rather, circle, along which languages move diachronically is outlined by Haudricourt (1968). Without making reference to Haudricourt, Nettle (1999: 142-143) adds some flesh to the abstract proposal: Simple inventories develop by way of underarticulation where the communicative context permits, and complexification "through a combination of coarticulation and word truncation." When inventories become large and distinctive segments are closer together in the articulartory space, failure to distinguish adjacent segments may occur, again reducing the size of the inventory.

[38] Another such self-organization tendency in phonology is that of feature economy leading to a tendency to maximally exploit features for distinctive purposes. Feature economy typically pertains to features that are lexically distinctive (Clements 2003: 328), and thus is a phenomenon interacting with the lexicon.

ing the so-called equi-complexity axiom for instance uttered by Hockett (1958) but said to have still earlier precursors (Kortmann and Szmrecsanyi 2009: 266), according to which all languages have overall the same degree of "complexity" (how this should be defined precisely is a matter of debate and there is no apparent consensus in the literature. Miestamo 2008 distinguishes two basic readings: the absolute one, where complexity is taken as an objective measure characterizing the linguistic system, and the relative one, where complexity is equaled to cost or difficulty for language users, both native and non-native; see also Nichols 2009a for suggestions), with "complexity" in one area of the grammar (say, morphology) balanced by simplicity in another (say, syntax). Nichols (2009a), drawing on a typological sample, reports having found neither evidence for a preferred level of linguistic complexity in her metric, nor for a functional trade-off between complexity and simplicity in different sub-domains of grammar. The correlations found here can, if one wants, be construed as evidence not for such a functional trade-off in complexity between different subsystems of grammar, but between phonology, root shape, and lexicon, at least if one is willing to equate morphological complexity in analyzable lexical items with "complexity" in one of the senses used in the recent literature, and if one agrees, which is perhaps less controversial, to calling languages with small consonant inventory systems, simpler syllable structure, and shorter lexical roots, more "simple" morphophonologically than languages with the opposite properties.

But first, before accepting either the narrow or the broad explanatory framework for the correlations with the morphophonological factors as explanatory, it is of course necessary to consider possible alternative hypotheses, and to see whether any fares better. This entails looking at both other correlations with structural factors emerging from the preliminary tests based on WALS, as well as other other possible explanations that come to mind, to see whether there are serious alternative explanations available.

5.4.2.11. Other significant correlations with WALS

In § 5.4.2.1., preliminary tests suggested interactions with a number of WALS features, of which so far only those pertaining to phonology have been discussed in greater detail. As for the other features, Mixed Models taking into account areal factors lended no support to an interaction with the order of adjective and noun (p = .086) and purpose clauses (p = .244). The remain features "survived" this additional control, and will be discussed in the following.

There is a very significant interaction when testing for correlations with the WALS features with the presence of possessive classification and the size of the classes in such systems when they are present. This feature remains significant when controlling for area in a Mixed Effects Model and the factor itself has significant power to predict the degree of analyzable items at p = .0004. There is an overlap of twenty-seven languages between the two samples on which the statistical test is performed; however, only two of these, Khoekhoe and Kolyma Yukaghir, are spoken outside the Americas. As noted by Nichols and Bickel (2005c), most often such systems are binary, in which case it is more widely known as a contrast between alienable and inalienable possession. Rather than seeing this phenomenon as being primarily driven by the semantics of the possessed element (for instance, kinship terms and body-part terms are semantic fields that are fre-

quently inalienably possessed), they conceive of it as being primarily lexically conditioned. Figure 30 shows that the more possessive classes there are, the higher the number of analyzable terms among those investigated.

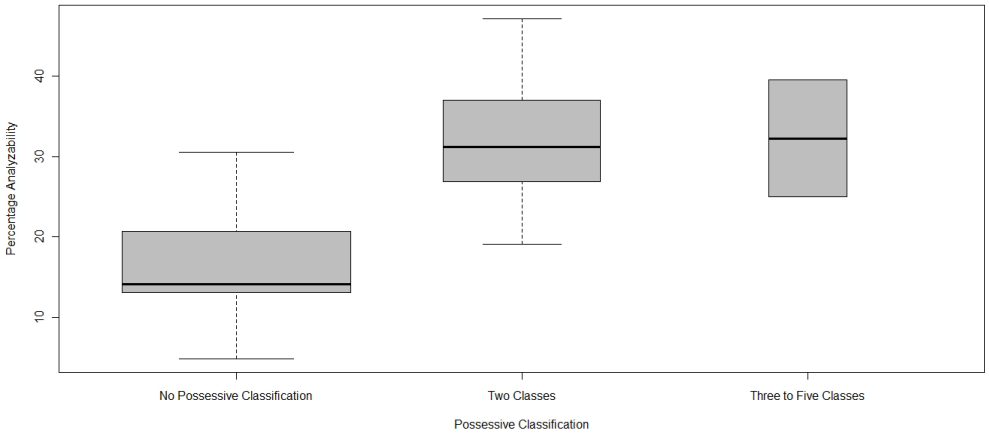


fig. 30: analyzable terms depending on possessive classification

Indeed, there is some evidence that possessive classification may be used to bring to light by morphological means different aspects of the semantics of lexical items. This point is made by Aikhenvald (2007: 38), who notes that in Tariana, an Arawak language, the same lexical item, *kare*, means 'wind' when alienably possessed and 'my breath, my heart' when inalienably possessed and prefixed with the respective marker *nu-*. However, there is little evidence from the sample data that this distinction is exploited on a larger scale to enrich the lexicon, although it is to some extent in languages of the Americas, as in Tariana. One sample language where it appears to be exploited to some degree is San Mateo del Mar Huave. Here, inalienable possession is marked by the suffix *-aran*, and the semantic domains it applies to mostly are, as is typical, body-part and kinship terms. Inalienable possession is optionally marked by the prefix *mi-* (Stairs and de Stairs 1981: 291-292, the authors do not use the terms inalienable and alienable possession, but it seems clear from their discussion that this is a typical system of possessive classification).[39] Often there is no apparent effect of the suffix for inalienable possession on the semantics of the root, as in (4.).

(4.) *mijiw-aran* 'breast/teat-INAL.POSS'

However, at times, roots bearing the suffix differ from those without it semantically, and it seems that it is used in a derivational fashion in (5.).

[39] Stairs and de Stairs (1981: 294) note that *-aran* can be added to a nominalised verb in which case the new form conveys 'clasificación,' as in *ajiüng* 'pray' – *najiüngaran* 'prayer.'

(5.) a. *mipeparan* /mi-apep-aran/ 'AL.POSS-inflate/globe-INAL.POSS' = 'bladder'
b. *omeaats-aran* 'inside-INAL.POSS' = 'heart'

A further possible example are terms for the 'testicles' in Ineseño Chumash, which consist of words for 'pit, seed' and 'stone' with the possessive prefix *is-*.

However, analysis of the data in the validation sample does not lend support to this evidence. Quite to the contrary, the estimate between no possessive classification at all and two classes is negative as opposed to positive here (-7.150) and thus not at all within that of the original sample (14.685 ± 3.134). The same is true for the features dealing with predicative adjectives.

As for the order of demonstrative and noun as a possible predictor, which is visualized in figure 31, validating the results of the original sample is difficult, because the languages in the validation sample fall in two groups only, those with demonstrative-noun order and noun-demonstrative order, with none of the rarer types involving demonstrative affixes and others with demonstrative elements on both sides of the noun mixed behavior figuring in this sample.

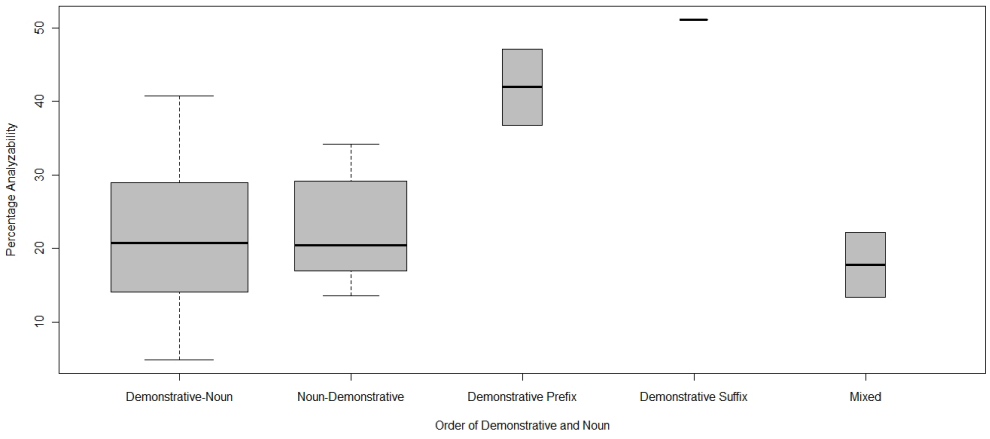


fig. 31: analyzable terms depending on the order of demonstrative and noun

As for the difference between the first mentioned major types, results are similar (1.254 ± 2.367 vs. .175), but since the drastic differences causing the original model to become significant occur with the types involving affixed (estimates are 20.367 for demonstrative prefixes and even 29.517 for demonstrative suffixes), the similarity between the results regarding the major groups is not very informative. In the original sample, languages with demonstrative prefixes are Abzakh Adyghe and Pawnee, and the one language with suffixes is Kiliwa, which all have unusually high percentages of analyzable terms. Given that there are thus only three relevant observations available, further data for languages with demonstrative affixes would be required to give a definite answer to the question whether this factor influences the behavior of languages with regard to analyzability. For now, the result is suspicious of being purely accidental.

As for semantic distinctions of evidentiality, the correlation shown in figure 32 remained significant in a Mixed Model taking into account areal factors at p = .0069, and it was possible to replicate the difference between languages with no grammatical evidentials and those with indirect evidentials, but not that between the latter and those featuring also direct evidentials (5.812 ± 3.143 vs. 6.162 and -8.015 ± 4.019 vs. 6.462).

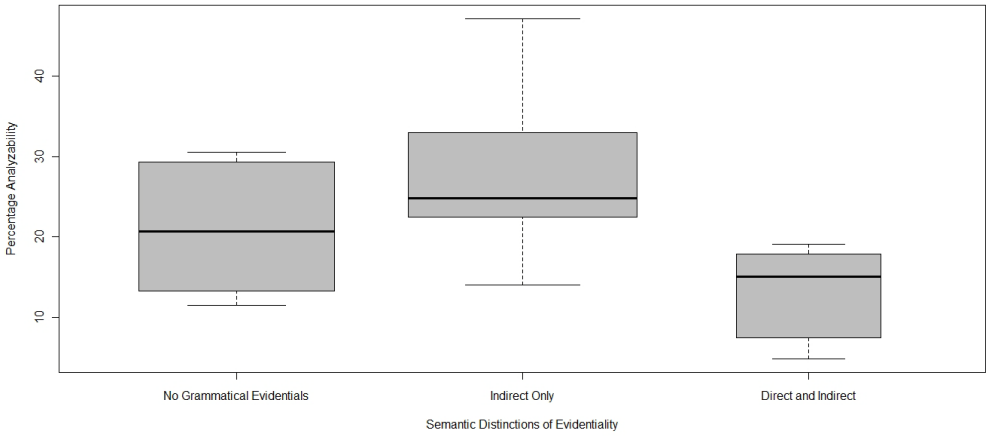


fig. 32: analyzability depending on semantic distinctions of evidentiality

Given this mixed result and the fact that it is unclear why the correlation should be there in the first place, the conclusion is that semantic distinctions of evidentiality do not seem to influence the number of analyzable lexical items to the same degree as the phonological features do, although further testing would be required to ultimately rule out a true effect.

The interation with predicative adjectives remains significant at p = .03823 when controlling for area, but the effect cannot be replicated on the dataset of the validation sample (estimates -6.539 ± 3.438 as opposed to -17.956 for nonverbal encoding and 8.836 ± 5.5 as opposed to 2.019 for mixed encoding), which suggests that the effect is not genuine.

Summing up, there is little evidence that among the structural features coded in WALS, any other than the phonological ones play a role in shaping the degree to which languages resort to analyzable lexical items.

5.4.2.12. Further tests and possible factors

5.4.2.12.1. Sociolinguistic Function: Esoteric vs. Exoteric Languages. Thurston (1989) proposes a distinction between languages with respect to their sociolinguistic function that is said to correlate with structures in grammar and in particular the lexicon. Esoteric languages, according to Thurston (1989: 556), "function primarily as codes for communication among people of the same social group. Over time, they tend to become gradually more complex. That is, they acquire a relatively high degree of allophony and allomorphy; they build large vocabularies with many near-synonyms and many opaque idioms; and they come to make relatively more numerous obligatory grammatical distinctions." Exoteric languages, in contrast, "have, as at least one of their primary sociolinguistic functions, use as a lingua

franca between peoples of different social groups. They tend to be structurally simpler than esoteric languages, because they must be easily learned by adults with different linguistic backgrounds" (Thurston 1989: 557).[40] More specifically, Thurston (1989: 567) argues that one diagnostic for exoteric languages may be "the relative lack of monomorphemic lexemes, particularly for terms that are usually considered endolexical [i.e. terms belonging to basic vocabulary]." This is, according to him, the result of coinage by second language learners when the name in the target language for a specific extralinguistic entity is lacking. Thurston states that "[w]hen more data of this sort are collected, I anticipate that a correlation will be found between the degree of esoterogeny and the number of highly specific monomorphemic lexemes."

This anticipation is open to empirical investigation using the data of the present study. In order to test Thurston's prediction against the present data, information on the sociolinguistic function of the languages in the statistics sample was gathered. In particular, attention was paid to whether the languages in the sample do have second language learners (however many) or not. These data were obtained primarily from the consulted sources for each language themselves or from Lewis (2009) and are found in Appendix C. This is a rather coarse measure, and it is acknowledged that it simplifies Thurston's more complex scenario somewhat in order to make it testable empirically. On the other hand, the coding should mirror Thurston's distinction to a reasonable degree, since exoteric speciation in his sense necessarily entails second language learners while esoteric speciation does not.

After constructing a Mixed Model design as usual, there was no appreciable difference between languages of either kind in the presence of morphologically complex terms itself (p = 0.7044). As the estimate of the model at 1.219 shows, it is even the case that languages without L2 learners have a slightly elevated number of morphologically complex terms when compared with languages that are not learned by second language learners, but even this observation is clearly not strong enough to be of significance. A visualization of the values is in figure 33.

[40] For a recent application of Thurston's dichotomy from a cross-linguistic point of view, see Lupyan and Dale (2010). Wray and Grace (2005) also heavily borrow from Thurston's work, although they speak of exoteric vs. esoteric functions of languages rather then exoteric vs. esoteric languages themselves, correctly pointing out that one and the same language may be used both for in-group communication as well as communication with outsiders.

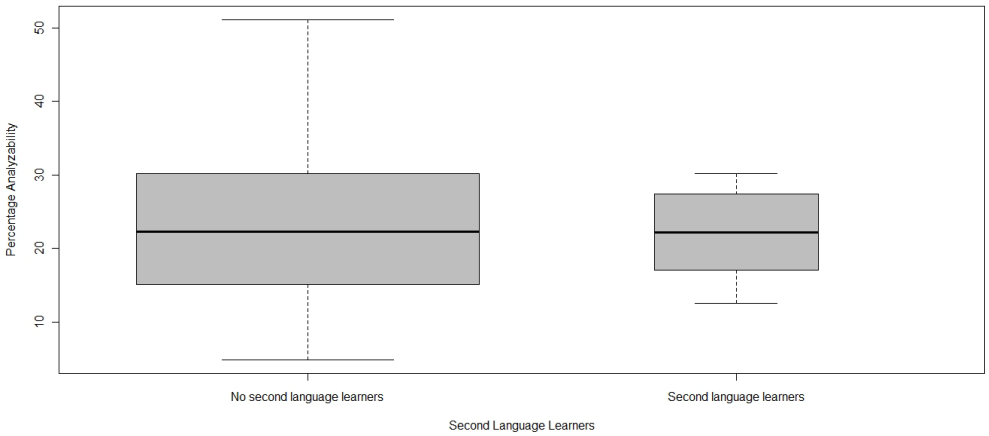


fig. 33: Morphologically complex items in languages with and without second-language learners

Nor was there a correlation with most of the other major global variables coded for each language (motivated terms in general, degree of metaphor and degree of contiguity, degree of deverbal formations).

Thus the data of the present study do not support of Thurston's expectation to find an elevated degree of complex terms in core vocabulary in exoteric languages (operationalized here as languages with second language learners). In fact, there is at least one case of a language outside the sample with sufficient published material on the matter where Thurston's predictions as to a correlation between esoteric speciation and monomorphemic lexical items do not go through. Yélî Dnye would be a textbook example of an esoteric language: it is spoken on an isolated island of the Pacific, and contact with outsiders is rare. And indeed, its grammar exhibits an enormous amount of complexity, as manifested in an elaborate apparatus of cross-referencing, often in a portmanteau fashion, with rampant morphophonological alternations, a highly suppletive verbal lexicon, and the probably most complex phonological system in the Pacific, featuring a number of cross-linguistically extremely rare sounds. This exuberant complexity causes that the language is rarely if ever successfully learned as a second language, and even women from other islands who marry a Rossel islander usually learn the language only very imperfectly (Levinson 2006a: 20-21). Still, Levinson (2006b: 230) notes that "Yélî Dnye is a language where many important, commonly employed nominal concepts are expressed with compounds." Judging from the evidence (also for body-part terms, the domain from which Thurston's original examples come) presented in Levinson (2006b), the language is not very different in this regard from Thurston's example of Anem, a exoteric language in his terms, adduced as support for the language's exoteric speciation.

On a related note, tests using the number of speakers of the languages in the statistics sample as a predictor variable for any of the major variables surveyed were carried out, with no significant results (the *p*-value for the percentage of analyzable terms is .3074

and that for the percentage of derived terms after logarithmic transformation is .2). In particular, a correlation between word length and presence of language contact and concomitant bilingualism as suggested by Trudgill (1996) on the basis of differences in word length between Standard Greek and northern dialects, which are in contact with neighboring Balkan languages on the basis of the first 50 items of the Swadesh list, could not be found on the basis of the present data on the language rather than dialect level ($W = 239.5$, $p = .6635$, Wilcoxon rank sum test).

5.4.2.12.2. (Large-Scale) Borrowing. As is obvious, heavy borrowing has the potential to have profound effects on the degree of analyzability of the lexicon, simply by the fact that in borrowed words (at least when defined strictly as the transfer of lexical material, that is, excluding calquing), possible internal morphological structure in the donor language is lost in the recipient language. Of course, derivational morphology may also be borrowed along with lexical items and subsequently nativized, as has been the case for instance in English borrowing from French, but it is probably safe to say that in most instances of borrowing, this is done at the expense of possible internal structure in the donor language. For instance, Sasse (2001: 503) appeals to the long history of mutual borrowing in languages of Europe to account for the "inexhaustible number" of simplex lexical stems found there. When it comes to large-scale borrowing, Australia also immediately comes to mind. Dixon (2001, 2002) proposes that for the most part of the continent, neighboring languages, due to extensive bilingualism teaming up with avoidance registers causing a constant need for replacement vocabulary for taboo words, on the long run end up sharing about 50 per cent of vocabulary, irrespective of genetic relatedness. This is known as the "50 per cent equilibrium model." An important study on this is Heath (1981), describing the situation in languages of Western Arnhem Land. However, this model is not universally accepted by Australianists, and the effects of word taboo are said to be overestimated by Alpher and Nash (1999). Evans (2005), while admitting that it is possible to reach figures as high as 50 per cent of shared vocabulary, adduces evidence from several Australian languages in contact, but still with undramatic levels of shared vocabulary. His conclusion is that there is significant variation in Australia in the extent of borrowing from area to area, depending also on the nature of social relations between speakers, and that the 50 per cent equilibrium model is not empirically well-substantiated on a larger basis in the Australian area.

Now, coming to the relevance of this in the present study, it is the case that two sampled Australian languages, Ngaanyatjarra and Nunggubuyu, have extraordinarily low numbers of analyzable terms, while a third, Gurindji, has somewhat more, but is shown by McConvell (2009b: 794) to feature many loanwords, around 45 per cent of all items in the World Loanword database and thus drawing close to the 50% figure Dixon's model. However, body-part terms are only moderately often borrowed. There is also one clear, instance of taboo-induced replacement, but McConvell (2009b: 797) denies strong effects of taboo on borrowing in Gurindji as proposed by Dixon.

Borrowing may also well play a role in producing the small number of analyzable terms in Ngaanyatjarra and Nunggubuyu, but another fact about these languages is that, like in many other languages of Australia, lexical roots are quite long in terms of number

of syllables. Yir Yoront, in contrast, which, unusually for an Australian language, features productive compounding, and where, interestingly and equally unusually for Australia, roots are generally short, monosyllabic or at the very least disyllabic, has the highest percentage of analyzable terms of the sampled Australian languages. This fact opens up a more parsimonious explanation, also in light of the controversiality of the status borrowing has in languages of Australia in the theoretical discussions, in that a cross-linguistically valid tendency, namely for languages with long lexical roots to have fewer analyzable terms than those with shorter roots, can be used to explain the differences in analyzability in the sampled Australian languages. This is simply the application of Ockham's razor, and it is not claimed that borrowing has no role to play, both in Australia as well as in the rest of the world. Indeed, data from the World Loanword Database can again be adduced to assess the question of interrelations between borrowing behavior and morphological complexity. Using the same subset of languages as in § 5.4.2.7.1., there indeed is a correlation at p = .02532 between the number of analyzable and borrowed terms to the effect that where there are many borrowed terms in the language, the simplicity score is higher, as seen in the plot in figure 34.

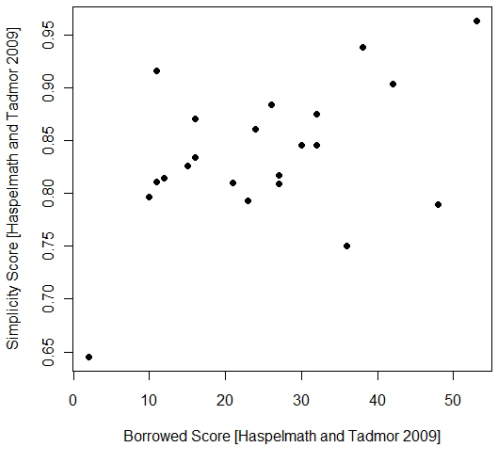


fig. 34: Correlation between borrowed and simplicity score in a subset of languages of the World Loanword Database

However, the positive correlation crucially hinges on the behavior of just one language, Mandarin Chinese, which is represented by the dot in the lower left corner in the plot in figure 34. This is relevant because for this language in particular, the differences between this study and the World Loanword Database in the assessment of analyzability is an important factor, in that complex terms of the redundant type are not motivated as defined in § 3.6.1., and thus not counted here, while they are in the World Loanword Database. If the peculiar case of Mandarin Chinese is removed from the dataset, the correlation also ceases to be significant (p = .159).

What is more, if the preliminary evidence from the case study of the Americas, which suggests that the predilection of a language for borrowing is not entirely independent of its lexical profile with respect to analyzability in native vocabulary (see § 5.4.2.7.1. for details), is valid, it enables one to treat borrowing behavior as a result of lexical organization with respect to analyzability rather than its cause.

5.4.2.12.3. Word Taboo. Taboos against naming the dead appear to be widespread around the world. Kroeber (1925/1976: 360) says that this principle is widespread in California, and, as a case in the American Northwest, Elmendorf (1951: 207) argues that in the Salishan language Tswana "the spread of derivative or compound descriptive terms through the lexicon, these terms originating as coined substitutes for tabooed words" is a likely concomitant of taboos against naming the dead, "an occasional but active custom." Since the taboo required words resembling the name of the deceased person not to be uttered, Elmendorf (1951: 207) concludes that in the course of time, the procedure would oust all words resembling a personal name. Comrie (2000) reports that in Haruai society people's names are identical to the names of everyday objects, and that at the same time a taboo against uttering the name of taboo kin is in place, which obviously leads to practical complications. For this reason, a large number of synonyms (many of them loanwords) exist in the Haruai language as a kind of backup for the event that the indigenous word should become unavailable. Of course, in a comparable situation where, unlike in the case of Haruai, neologisms are coined or related words are semantically extended rather than words borrowed from foreign languages, this would lead to a notable increase in lexical motivation. Indeed this is apparently the case in a large number of Austronesian Languages, where people's names also coincide with lexical items (Simons 1982, see also Rensch 2002: 192-195 for brief discussion of Polynesian specifically). It would be extremely interesting also to ascertain in which parts of the world people's names are at the same time ordinary words in the language with lexical meaning.[41] Further, Gamkrelidze and Ivanov (1995) attribute a large number of lexical innovations in different branches of Indo-European to taboo-induced replacement, most often in the case of names for animals which are said to have ritual significance (see also Emeneau 1948 for discussion of "hunter's taboo," forbidding to utter the name of an animal being hunted, which also seems to be very widespread cross-culturally), Tetun features a special register called *lia tasi* used while at sea fishing (van Klinken 1999: 8-10), and Barlaan (2003) discusses replacement vocabulary during the rice-harvesting season among the Isnag.

The extent of the phenomenon and its precise characteristics in different parts of the world is unfortunately not at all clear. One clear case for the presence of different subtypes of naming taboo on an entire continent is Australia (Dixon 2002), although the usually assumed far-reaching effect of this cultural practice on the lexicon of Australian languages has been challenged (Alpher and Nash 1999).

[41] This is frequently indicated for Buin in the consulted source: for instance, *kuruku* 'thunder' is a female name, while, to adduce data from a language from another area of the world, Nez Perce *símux* 'charcoal' is also indicated to be a man's name. Thus, the data of the present study indicate that the phenomenon is well attested, but are not sufficient to allow for more systematic exploration.

Testing effects of word tabooing cross-culturally is not an easy task, because no large-scale comparative anthropological treatment of patters of word taboo is presently available that would make clear just in which cultures word taboo rules are in place and where such practices are unheard of. Thus, there may be instances of erstwhile complex taboo words being conventionalized in the ordinary lexicon (and in Tswana precisely this seems to be the case to some extent). For instance, Koyraboro Senni has *taa-haa* 'sew-thing' for 'needle' to replace the monomorphemic ordinary term *sana* which must not be used at night. The question is how pervasive influence of word taboo can be on the nominal vocabulary as a whole, and whether it is strong enough to be capable of shaping lexical structures on a large scale rather than replacing single lexical items every now and then in a piecemeal fashion. This remains unclear. For the time being, it is possible to at least note that the most widespread case of taboo words reported in the literature pertain to the names of (predatory or game) animals and are thus unlikely to influence the percentage of overt marking in the meanings investigated here. Further, there are languages in the sample with both a very low degree of analyzability in the lexicon where there is no evidence for any sort of word taboo being operative. For instance, for Bora, a language with comparably many morphologically complex terms in the lexical items investigated here, Frank Seifart (p.c.) reports that any practice of word tabooing is unknown to him. Likewise, Zaira Khalilova and Madzhid Khalilov (p.c.) report no evidence for practices of word taboo in Bezhta, a language with few analyzable terms. Explicit statements in the literature for the absence of word taboo are unsurprisingly rather hard to come by, but Epps (2008: 15) mentions that Hup society, speaking a language with a relatively high degree of analyzable terms is egalitarian and liberal, with few social taboos and restrictions. This shows at the very least that presence or absence of word taboo cannot be the single underlying cause of differences in analyzability on the lexicon. Also, the outcome of taboos may be quite different: either it can lead to (massive) borrowing, as stated for Australia with its widespread use of replacement registers for certain kin relations, or it can lead to descriptive neologisms to replace the tabooed lexical item, so that it could influence the lexicon theoretically in either way with respect to the degree of analyzability.

5.4.2.12.4. Syntheticization. An explanation of increasing morphological complexity in the lexicon not directly related to phonological factors, but appealing to learning difficulties of great allomorphic variation is offered by Fortescue (1992) for polysynthetic languages[42] in general, using Eskimo-Aleut as his example.[43] Fortescue (1998: 49) summarizes that

[42] Fortescue is aware of the difficulties in defining polysynthesis (cf. also § 4.5.1.2.2.), and says that polysynthetic languages have traits that in sum allow "the expression within complex word forms of numerous elements that in more analytic languages correspond to independent lexical items, verbs thus often corresponding to whole sentences in the latter"(Fortescue 1992: 242fn1).

[43] Proto-Eskimo is reconstructed by Fortescue et al. (1994: xi) as having fifteen native consonant phonemes, with many additional non-native ones due to borrowing. It allowed maximally for CVC syllables. Canonical stems have (C)Vt(ə)- and (C)V(C)CV(C)- shape (and there are possibly also corresponding trisyllables) with some phonotactic restrictions as to what consonant may appear word-finally (Bergsland 1986: 98). This is not dramatically different from Central Yup'ik, the Eskimo-Aleut language in the sample, so that diachronic phonological pressure indeed does not seem to play a major role.

"Proto-[Eskimo-Aleut] must have lost a large portion of its previous stock of lexical items as capitalisation on its highly productive derivational apparatus increased and lexical gaps were filled more and more by derived forms from relatively few stems." Indeed, there are some 200 basic postbases (see § 4.4.2. for this term) in both Greenlandic and Central Yup'ik, with somewhat fewer in Aleut; 50 Aleut postbases have Eskimo cognates (Bergsland 1986: 102), suggesting an expansion of derivational postbases in the latter. Fortescue (1992: 245) argues that an expanding derivational apparatus with concomitantly increasing allomorphy creates an increase in memory load for the acquisition of its properties, and, once the process has begun, feeds into the development of polysynthesis, the outcome of which is a "reorganized state of balance between the inventory of lexical stems as opposed to productive bound affixes." Furthermore, he (1992: 246) states that "typically, polysynthetic languages do display a relative paucity of lexical stems, this being counterbalanced by an enormously increased derivational potential compared to more analytic languages." It would indeed be of great value to assess this impressionistic statement quantitatively on the basis of a sample of languages with a high degree of synthesis. But even with more systematic evidence pending, Fortescue's account has some merit in that it could explain a high degree of analyzable terms in many "polysynthetic" languages, in spite of a universally accepted definition still lacking. Note, however, that languages regarded as polysynthetic are not necessarily characterized by a rich derivational apparatus, Ket being an example of such a language (cf. § 4.5.2.1.). Furthermore, Fortescue's proposal would also fail to account for the behavior of languages with an isolating profile, such as Efik, Bororo, and Hawaiian. This should not be taken to mean that his proposal as to a shrunk inventory of lexical elements at the expense of increasing derivational possibilities is incorrect, but merely that it cannot account for all cross-linguistic variation with respect to differences in the percentages of analyzability, since, if indeed syntheticity were the sole responsible parameter, one would expect only such languages with a profile of lexicon-grammar-interaction as outlined by Fortescue for Eskimo-Aleut to be characterized by a largely analyzable lexicon, and not others. The following section discusses grammatical properties as a potential factor, with particular reference to a typical ingredient of polysynthesis: head-marking.

5.4.2.12.5. Other Grammatical Factors in the Distribution of Morphological Complexity?

Thanks to the work of Nichols (e.g. Nichols 1992, 1998, Bickel and Nichols 2009, in press), it is well-known that there is a world-wide cline in the distribution of certain grammatical features, such as head- vs. depending marking, inclusive/exclusive distinction in pronouns, numeral classifiers, as well as consonantism in pronominal roots (Nichols and Peterson 1996).

Could it be possible that there may also be grammatical factors that shape a language's behavior with respect to analyzability in its lexicon? Consider, for instance, the following examples of basic transitive constructions from Kiowa and Biloxi, which are typical for languages of North America.

(6.) a. *kɬyą́·hî· tʰàlí· ę́-góp*
man boy 3SG/AGT:DU/OBJ-hit/PF[44]
'The man hit the two boys' (Watkins 1984: 205)

b. *tohoxka ayeki duti na*
horse corn he.eats.it
'The horse eats the corn' (Einaudi 1974: 166)

These are head-marking constructions (Nichols 1986): arguments carry no markers indicating their grammatical function in the clause; rather, these are identified by means of affixes on the verb. In contrast, languages of Eurasia are predominantly dependent marking, as illustrated by the sample languages Bezhta:

(7.) *gedi ãq'o boxx-iyo*
cat.ERG mouse catch-PST.W[45]
'the cat caught the mouse' (adapted from Xalilov 1995: 410)

However, not all languages of Eurasia follow the typically dependent-marking clause alignment in this area. Two notable exceptions are the Yeniseian and Munda language families, as illustrated by Ket and Sora examples in (8.).

(8.) a. *hīɣ qímdɨ̈l dítòŋ*
hīɣ qímdɨ̈l du^{8}-i^{6}-t^{5}-a^{4}-oŋ^{0}
man girl 3M.SJ^{8}-3F.O^{6}-SU^{5}-D^{4}-see^{0}[46]
'The man sees the girl' (Vajda 2004b: 22)

b. *ənlen daʔa-n a- tiy- t- ay*
we water-N.SFX 1PL-give-NPST-1[47]
'We give (him/her) water' (Anderson and Harrison 2008: 328)

Sora in fact features the object marker *a'dəŋ* occurring in connection with lexical rather than pronominal arguments. It is grammaticalized from the possessed form of a word meaning 'body,' and is probably a recent innovation, as suggested by the fact that it is an independent word rather than an affix and restricted to animates, a semantic restriction that is typical for early stages in the grammaticalization of case markers (Hopper and Traugott 2002). Other major Munda languages, e.g. Mundari and Santali, lack marking of core arguments altogether, making Sora an unusual Munda language in this respect.

[44] Glosses: AGT 'agent,' DU 'dual,' OBJ 'object,' PF 'perfective.'
[45] additional gloss: PST.W 'witnessed past.'
[46] glosses: M 'masculine class (a subset of animate class),' SJ 'verb-internal subject agreement affix, or subject pronoun,' F 'feminine class (a subset of animate class), O 'verb-internal direct object agreement affix, or direct-object pronoun,' SU 'suppressive adposition (verb affix denoting superficial contact with an object), D 'durative marker (appears in many stative and activity verbs).'
[47] additional gloss: NPST 'non-past.'

Crucially, not all languages of the Americas are head-marking. One example of a strictly dependent-marking North American language is Wappo, formerly spoken in California (Wappo is a so-called marked nominative language, but this does not affect its characterization as being dependent marking), as seen in (9.).

(9.) *ce k'ew-i ce holo:wik'a t'a-ta?*
DEM man-NOM DEM snake kill-PST[48]
'the man killed the snake' (Thompson et al. 2006: 11)

The point of discussing these examples is that Ket and Sora, typologically unusual languages for Eurasia, receive after Abzakh Adyghe (which is a double-marking language on the level of the clause in terms of Nichols 1986, but has many head-marking traits) the highest scores in the degree of analyzability in the nominal lexicon, while Wappo, a typologically unusual language for North America overall, receives the lowest score in analyzability of all North American languages in the statistics samples. This suggests that there are other structural features, aside from phonology, in play when it comes to the shape of the nominal lexicon. In § 4.6.5.4., it was suggested that head-marking elements in Kiliwa may be a factor facilitating the coinage and conventionalization of complex clausal nominals. While head- as opposed to dependent-marking is a typological factor that may be applied on different levels of linguistic structure, including morphological marking within the noun phrase as well as clause-level and even interclausal syntax (and these patterns may be used to jointly define a profile of individual languages with respect to the parameter, Nichols and Bickel 2005b), the focus will here be on the level of the clause, at the expense in particular of marking in the noun phrase. This is not to say that NP-level marking would not be interesting to investigate.

The question whether there are differences in the lexicon depending on preferred marking patterns on the clause level is tested in the following fashion: rather than using one overall metric assigning languages to one type (Nichols and Bickel 2005a), the data on verbal person marking from Siewierska (2005), with amendments to fill gaps in the data as already used in § 4.7., provides one measure of indexing on the verb (=head-marking). In addition, data were gathered for the sample languages on whether core grammatical relations are flagged by case markers or case-like elements such as adpositions (=dependent-marking). Data are in Appendix C. This in effect creates two independent parameters for head- and depending marking elements (cf. Cysouw 2002): a dependent-marking language on the clause level is defined as one with core cases but no verbal person marking, while a language with dominant head-marking elements on the clause level is one with no core cases, but verbal person marking for both A and P arguments. As tables 19 and 20 show, the properties are areally unevenly distributed under the Dryer-6 breakdown (cf. also Nichols 1992). The differences are significant at $p < .001$ for verbal person marking and at $p < .02$ for presence vs. absence of core cases by Fisher's exact tests, so statistical modeling once again needs to take these differences into account.

[48] Glosses: DEM 'demonstrative,' NOM 'nominative case,' PST 'past tense.'

	Africa	Australia-New Guinea	Eurasia	North America	South America	Southeast Asia and Oceania
No verbal person marking	2	2	3	1	3	5
Only A	1	1	4	1	4	0
A and P	0	2	4	15	8	0

table 19: areal breakdown of types of verbal person marking

	Africa	Australia-New Guinea	Eurasia	North America	South America	Southeast Asia and Oceania
Languages with core cases	2	7	9	5	12	1
Languages without core cases	2	3	2	16	13	5

table 20: areal breakdown of presence vs. absence of case marking for core grammatical relations

Contrary to the hypothesis generated by looking at the languages mentioned above, no clear impact of differences in the locus of marking on the clause-level was revealed by a Mixed Model controlling for area emerged, neither for the combination of the two variables of verbal person marking and core cases (p = .4476), nor for one of them separately (person marking: p = .6484 , core cases: p = .3134).

As observed by Nichols and Bickel (2005b), genetically related languages sometimes differ in their marking type. In particular, they note that within Uto-Aztecan, Pipil is a consistently head-marking language, without cases for core grammatical relations but with affixes on the verb cross-referencing arguments (Campbell 1985: 39-56; 74), while Tümpisa Shoshone is consistently dependent marking, featuring a nominative-accusative case system and no indexing of arguments on the verb (Dayley 1989a: 53-54; 176-178).

To see whether there is any noticeable impact of these differences, Tümpisa Shoshone equivalents for the full 160-meaning list were gathered from Dayley (1989b), yielding 126 of 160 possible equivalents (this was done after most calculations were computed and data for chapter 6 were collected, so the Tümpisa Shoshone data is not otherwise evaluated systematically). The result is that 29 per cent of these were analyzable (as opposed to 18.2 in Pipil), with 34.3 per cent being of the derived type (as opposed to 25.9 per cent in Pipil). According to the hypothesis, Pipil should have a higher number of analyzable terms than Tümpisa Shoshone, but it does not, thus showing that any immediately effects of locus of marking in the clause on the degree of analyzability in the lexicon seems unlikely.

5.4.2.12.6. Effects of Mode of Subsistence on Analyzability in the Lexicon? The lexicon is probably the subsystem of language which is most directly influencable by non-linguistic factors, be they cultural or environmental. It is therefore conceivable that the lifestyle of a speech community will be a factor that influences the structure of language, as suggested e.g. in Brown (2005a, b) for certain features of the lexicon specifically (see also Cysouw and Comrie forthcoming for some possible grammatical correlates). Two different data sources are used to address whether there are such differences in the major quantitative variables concerning the lexicon surveyed in this work. Hammarström (2010, online appendix) provides data on the dominant mode of subsistence for the world's language families. Hammarström employs a binary classification into hunter-gatherers and agriculturalists. In addition, data from Murdock and White (1969) provide more detailed information on mode of subsistence for a selection of world cultures. On the basis of the information from Murdock, languages where grouped according to whether the main contribution to mode of subsistence is provided by (i) hunting and gathering, (ii) horticulturalism or pastoralism, or (iii) advanced agriculture. In cases where two of the above factors are said to contribute equally, the culture and its corresponding language was coded as belonging to the category with the lower number. For instance, cultures which rely on both hunting and gathering and horticulturalism or pastoralism were treated as hunter-gatherers. This is simply a measure to avoid ambiguities and thus to allow for statistical analysis. However, the overlap between them and the corresponding languages presently surveyed is rather small, which is why the data from Murdock (1969) were amended by extraction of relevant information from Levinson (1991).[49] Resulting data are in Appendix C.

Testing for effects on the degree of morphologically complex terms on the basis of both datasets for mode of subsistence using a Mixed Model controlling for areal effects reveals at best borderline significance for the Hammarström dataset (likelihood ratio test: *p*-value for factor: = .1010, estimate: 3.743) and no significance for the Murdock and White/Levinson dataset (likelihood ratio test: *p*-value for factor =.2335). Figures 35 and 36 plot the results.

[49] When data on a particular group are available in both sources, data from Murdock (1969) were used.

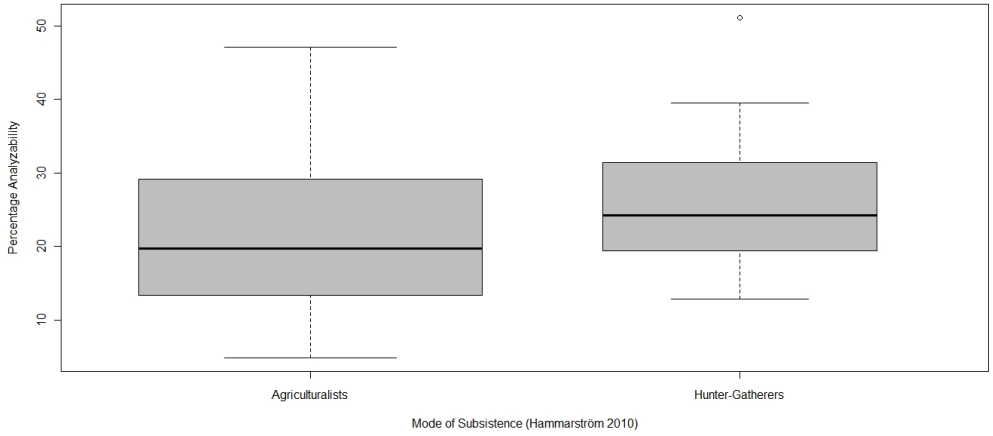


fig. 35: differences in the degree of analyzable terms between agricultural ists and hunter-gatherers, data from Hammarström (2010)

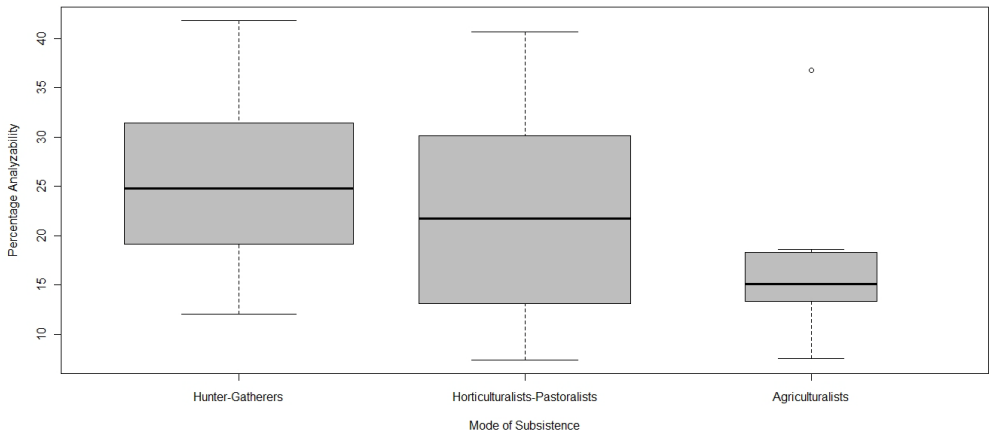


fig. 36: differences in the degree of analyzable terms between agriculturalists, horticulturalists/pastoralists and hunter-gatherers, data from Murdock and White (1969) amended by data from Levinson (1991)

In both cases, languages spoken by peoples relying dominantly on hunting and gathering as the primary mode of subsistence turn out to employ morphologically complex expression to a higher degree than agriculturalists, although there also is an areal signal in the data, in particular in the Murdock and White/Levinson data, and more importantly, the difference is insignificant. If there should be a genuine effect in spite of lack of clear statistical significance, it points to a further area of interaction between language and culture

which seems worth probing in more detail, although at present it is not clear just what should be the cause of the effect of mode of subsistence on the structure of the lexicon. Like Cysouw and Comrie (forthcoming), here, the discussion abstains from speculating about possible causes and restricts itself to simply reporting what can be observed.[50]

5.4.2.12.7. Creolization. One notable property of the behavior of the two creoles in the sample, Bislama and St. Lucian Creole French, is that it is remarkably unremarkable. This is in stark contrast with the commonly uttered opinion that "[p]idgins and creoles exhibit a high degree of motivation and transparency in compounding as a direct consequence of their small vocabulary" (Romaine 2002: 1094).[51] The evidence gathered for this study receives additional backup by the fact that the Creole language in the sample of Haspelmath and Tadmor (2009c), Seychelles Creole, receives the second lowest simplicity score of all languages in the sample exceeded in rarity of complex expressions only by Tarifiyt Berber (Bradley Taylor p.c.). If there is anything remarkable about the behavior of the creoles in these two studies, it is that they, quite contrary to what one might expect, come out rather at the lower end of the continuum on which languages are categorized with respect to the presence of analyzable terms. There are, judging from the evidence from both samples, a large number of non-creoles that regularly outreach the creoles with respect to the quantity of complex terms. This might be interpreted as being due to the fact that creoles, in the process of evolution from earlier pidgins (if they indeed evolved from pidgins, the notion that this always needs to be the case seems to be increasingly questioned), lexicalization (in the diachronic sense of univerbation) has rendered a large number of erstwhile compounds or circumlocutory phrases unanalyzable (a process alluded to by Romaine 1988, 2002: 1094 for Tok Pisin specifically). However, there is no evidence for such a development on a large scale in the data for Bislama and St. Lucia Creole French. Rather, the simplex lexical items in each language can be readily traced back to simplex lexical items in the respective lexifier language.

Likewise, referential expansion by means of polysemy as a technique to enrich expressive possibilities in creoles, as mentioned e.g. by Holm (1988: 108), is not present to a significantly higher degree when compared with the world-wide situation in non-creoles. On the basis of this study, it is not possible to confirm or refute claims such as that "there

[50] A further possible correlation that was tested is that between the occurrence of colexification of 'milk' and 'breast.' When one thinks about what pragmatic factors may give rise or maintain the colexification of the two referents, it seems obvious that it must be utterances in the context of nursing, such as "the baby wants [milk/breast]," where in fact reference is ambiguous (compare the following example sentence for Cashinahua *chuchu*: *Chuchu manuikiki. Amave.* 'He wants milk/breast. Let him drink.' (original translation: "Desea leche. Hazle tomar."). In contrast, in societies with an advanced mode of subsistence involving domestication of animals, in other words, where milk may be assumed to be a regular part of the diet, other contexts in which 'milk' occurs may in fact be more salient. However, the results were negative: there was no effect of mode of subsistence on this pattern of colexification when testing with both datasets on mode of subsistence.

[51] Relevant for the present study in general is also Rice's (to appear) description of the inventory of lexical stems in Athapascan languages as "staggeringly small," which is why according to her stems are "routinely called upon semantically to do double and triple duty, if not more, through conversion, compounding, juxtaposition, and inflection." Note that the analyzability score for the sampled Athapascan language Carrier is among the highest in Western North America.

is ... every indication that the lexicons of early (i.e. non-extended) pidgins are very much smaller than those of natural languages" (Holm 1988: 108), simply because the two languages in the sample are creoles and not pidgins in their early state of their development. But as far as creoles are concerned, from a cross-linguistic point of view, there is nothing special to be noted about them. However, statements emphasizing a high degree of motivation in pidgins and creoles do have some justification. When compared with their lexifier language, it may well be true that there indeed is a notable increase in the usage of compounds and polysemy, as can be illustrated by contrasting Bislama and English: a number of unmotivated simplex terms of English have not made their way into Bislama, their meanings being rendered by complex items, such as *ashes* vs. *sit blong faia* 'shit of fire,' *nest* vs. *bed blong pijin* 'bed of bird,' and many more. Likewise, *smok* in Bislama can not only mean 'smoke,' but also 'dust,' and *nus* does not only denote the 'nose' but also 'nasal mucus' and 'froth, foam.' But note that English, although not included in this study, is extremely likely to participate in the language area comprising Eurasia and Northern Africa with low degrees of complex lexical items and polysemes (evidence for the validity of this assumption is in Urban 2008). So while the degree of lexical motivation is elevated in creoles when compared to the largely unanalyzable stock of vocabulary items in the lexifier languages, which is, at least in the case of Indo-European-based pidgins and creoles unusually poor in motivated words, there seems to be no basis for the claim that creoles in general have elevated ratios of complex and polysemous lexemes when compared against the cross-linguistic situation. Rather, they seem to have, like their European lexifier languages, a comparably low degree of motivated terms, although somewhat higher than the languages they are descendent from.

As far as the specific semantic associations by means of compounding and the types of occurring semantic extensions of lexifier-language lexemes are concerned, substrate influence rather than creolization-specific universal processes appear to plays a significant role. For instance, the presence of a number of complex expressions on the basis of *sit* 'shit' in Bislama, one of which was mentioned above, appears to mimic structurally similar formations that seem to be common in Oceania as a whole (see § 6.2.3.3. for discussion of such extensions). For the case of Bislama specifically, Camden (1979) amasses evidence that the semantic structures in the lexicon (as well as in syntax) in particular match that of the Oceanic language Tangoa to a high degree, his conclusion being that "while the Bislama lexical structure looks basically English to a native speaker of English, it also looks basically Tangoan to a native speaker of Tangoan" (1979: 54). Similarly, semantic extensions of lexical items in Jamaican creole noted by Cassidy (1971: 216), such as the extension of the word for 'sun' to also mean 'day' and the ability of the word to 'water' to also refer to bodies of water such as 'river' or 'lake,' is frequent in normally transmitted languages globally, including African languages that form the substratum of Jamaican Creole. Holm (2000: 104), in discussing compounds in Nubi, an Arabic-based creole of Africa, mentioned by Heine (1982: 20), notes that "[s]uch compounds may have resulted from a universal strategy for expanding a pidgin vocabulary to fill lexical gaps, or they could represent calques on compounds in substrate languages." Similar evidence is presented in Parkvall (2000: 113-114), leading him to assume an agnostic position as to the source of lexical structures in Atlantic creoles as well. While the more or less anecdotal evidence

presented above does not rule out the possibility that there may indeed be mechanisms of lexical expansion by formation of morphologically complex expressions peculiar to the process of creolization only, nor that it may indeed have happened that semantic extensions occurred in the context of attempting communication in a setting with extremely little shared vocabulary between interlocutors, in the light of the ubiquity of most semantic structures found in creoles (and sometimes thought to be peculiar to creoles), substratum influence seems in many cases to provide a simpler and more parsimonious explanation for semantic structures in creoles (see already Huttar 1975, who arrives at similar conclusions, albeit on a somewhat different route). At any rate, the data in chapter 6 may be of use for creolists in formulating more fine-grained hypotheses as to the question of the origin of creole semantic structures.

5.4.2.12.8. Concluding remarks. Previous sections discussed alternatives to an explanation in terms of phonological complexity and root structure. It turned out that, although effects of some of them cannot be ruled out, accounts based on them would be less stringent than the one appealing to complexity of the word and of the sound system, either because (i) the phenomena in question are not universally applicable since they pertain to certain types of languages only, or (ii) mostly yielded negative or equivocal results when analyzed by means of statistics. In summary, structural pressure arising from complexity of the word and of the sound system can for the time being be said to be the most plausible candidate to shape analyzability in the lexicon (although further evidence not presently available on each of the topics may change this assessment), even though the mechanisms underlying it are not entirely clear in their details, and the particular interpretation suggested here is open to revision and refinement, with studies of homonymy in actual speech events being sparse as they are.

At any rate, the obtained correlations remain an empirical fact. As Dryer (2003: 120) remarks: "While I share the interest that others have in explaining crosslinguistic generalizations, there is a sense in which such generalizations are more valuable than the hypothesized explanations, since we can often have a much greater degree of confidence in the validity of the generalizations themselves than we can have in the explanations that have been hypothesized for them." Then, the discussion can be concluded with a dialogue from Orr (1962: 17) that seems appropriate:

R. Do you mind if I rest awhile? I've just found a monster floating about in my psychological orbit, and I feel a little uneasy.
O. What is it?
R. Synonymic-homonymics, an ugly brute!
O. Perhaps we had better stop for a bit.

As another alternative, perhaps, as suggested by David Gil (p.c.), it would also be worth thinking in the opposite direction: if languages favor complex terms, they can live with simple phonological inventories (cf. also the alternative explanation, though not generally accepted, for the developments in Mandarin Chinese discussed in § 5.4.2.3.2., which has it that the phonological system only began to shrink after the introduction of disyllabicity).

Instead of elaborating on these issues any further, the following discussion is concerned with the focussing on nominal referring expression in this study, seeking at least to hint at some interdependencies with the overall lexical organization of the language's lexicon in terms of the two parts of speech held by most linguists to be universal: nouns and verbs.

5.5. NOUNS AND VERBS

5.5.1. General Analysis of Vocabulary

Having established that there are languages in which analyzability is pervasive in the nominal lexicon, a question one can ask is whether this is due to a general difference in the prevalence of nominal as opposed to verbal encoding of referents. Relatedly, the question also pertains to the relative frequency of simplex noun and verbs in the lexicon, which may be relevant, because, if a paucity of simple unanalyzable nouns can be diagnosed for a particular language, then this would correlate with an elevated degree of analyzable nouns as a sort of "replacement vocabulary" to make up for the paucity of root nouns. To contextualize the investigation, it should be pointed out that highly divergent organizations of the nominal and verbal domain have been noted in the literature. Pawley (1993) reports that in Kalam, a language of New Guinea, verbs are a closed class with very few members and quite generic semantics which are conventionally combined in larger constructions to yield more specific semantic content, while nouns are in contrast much more numerous, although also here morphologically complex expressions are found. In contrast, Talmy (2000) highlights the deverbal character of the Atsugewi nominal lexicon. While Kalam thus makes do with a small restricted set of verbs, they are of such importance in Atsugewi that they are the basis for the formation of the other major part of speech, nouns. In this sense, the investigation takes up the rough typology of basic lexical types (i.e. noun-based vs. verb-based) outlined by Talmy (2000), which is with a different approach also addressed by Nichols and Nichols (2007).

Elucidating this question is not easy since what is needed is a representative sample of general vocabulary for all languages to be tested (Nichols and Nichols 2007 restrict themselves to a small list of glosses the equivalents of which they search for in their test languages of the Caucasus and the Pueblo languages of North America). Since a more general assessment that aims at looking at the vocabulary as a whole is very time-consuming, the present investigation is restricted to a small set of test cases, consisting of data for only four languages, and because of this restriction, the generalizations to be drawn can be nothing but extremely tentative.

Representative languages were selected more or less at random, except for the fact that obviously they were meant to define extreme points on the continuum of analyzability in the nominal domain. Since what is of interest here is the behavior of languages with a highly analyzable nominal vocabulary, two such languages, Kiliwa and Pawnee, were analyzed to allow for comparison. Badaga was chosen as a language representing the opposite type, with very few analyzable nouns, and Koyraboro Senni as a language that falls somewhere in between the extremes. An important criterion was that dictionaries

are sufficiently large and can thus assumed to be more or less comprehensive. Another important requirement for the selection of languages was that the consulted source provide clear information as to the part of speech of the headwords; another criterion was that there is a grammar available (written in the case of Koyraboro Senni, Pawnee, and Kiliwa by the same author as the lexical source) that identifies the morphosyntactic criteria that allow for distinguishing between nouns and verbs (Parks 1976, Mixco 1965, 2000, Heath 1999, Balakrishnan 1999). The methodology is simple: a random sample of the vocabulary was gathered by reading every tenth page of the Pawnee/Kiliwa/Koyraboro Senni-English section of the dictionaries, and, due to its larger size, every 20th page of the Badaga-English section of the dictionary, beginning on the first page (see Nettle 1995 for a similar approach for generating a random vocabulary sample from dictionaries). This avoids both biases from (fossilized) prefixes of a certain shape that cause a particular part of speech to begin with a certain segment and thus to cluster in a certain region of the dictionaries (note that, for instance, in Meyah, many nouns begin with /m/, Gravelle 2004: 104). Entries for native unanalyzable nouns and verbs (that is, disregarding clear loanwords) on the pages read were counted and, in the case of nouns, their meanings were recorded alongside. In Koyraboro Senni, many stems are ambiguous as to lexical category and can function as either nouns or verbs (Heath 1999: 96). Such stems were not counted as being either nominal or verbal and were simply ignored. The same goes for the fewer number of such cases in the other languages, such as Pawnee stems functioning as verbs but that may be used as nouns by suffixation of the nominal suffix *-uʔ*.

This yielded a sample of 101 Pawnee words, 68 Kiliwa words, 177 Koyraboro Senni words, and 145 Badaga words, coded for whether they are defined by language-internal criteria as nouns or verbs. Subsequently the number of nouns was divided by the number of verbs to obtain a measure called the NOUN/VERB-RATIO here. A high noun/verb-ratio indicates that simplex nouns are more frequent than simplex verbs, and a low ratio indicates the opposite situation: unanalyzable verbs outnumber unanalyzable nouns. Table 21 provides the values for the noun/verb-ratio from the dictionary sample along with the number of analyzable terms on the list of 160 meanings.

	Percentage Analyzability	Noun/Verb-ratio
Badaga	9.4	7.056
Koyraboro Senni	13.6	1.77
Pawnee	47.1	0.46
Kiliwa	51.1	0.66

table 21: noun/verb-ratio and percentage of analyzable terms for the four test languages

As the values already show, the language with the highest noun/verb-ratio and therefore with the largest number of unanalyzable nouns, Badaga, is also the one with the fewest analyzable terms on the 160-meaning list, while the two North American languages with pervasive analyzability in the nominal lexicon, have a very low number of simple nouns as opposed to verbs (cf. in this context also discussion of Kiliwa summarized from Mixco 1965, 2000 in § 4.6.4.2.1., where the "verbal" character of the language is emphasized).

The Spearman's rank correlation is very strong at -0.8, and it is easy to mentally fit a regression line. Although the results must be seen as being preliminary in nature, they cast doubt on Dixon's (2010: 305) as strong as casual claim that "[t]here are never as many simple verbs as there are nouns." A plot of the correlation is in figure 37.

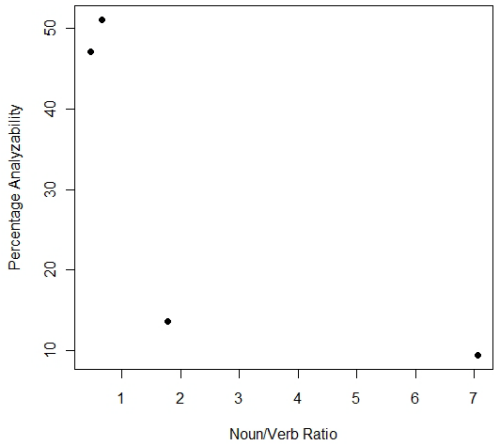


fig. 37: correlation between the noun/verb ratio and the percentage of analyzability

Another question that can be addressed with this data is: if the inventory of complex nominals in a language is very large and covers many meanings, what meanings, then, are expressed by simple nouns? This yields quite interesting results. As seen in table 22, unanalyzable nouns in Pawnee sampled from the dictionary are easily assigned to a small number of semantic domains: somewhat less than half of them are terms for animals and plants on the generic level. Other semantic domains in which unanalyzable nouns are found are kinship terms, body-part terms, topological and natural kind terms, and, frequently, names of tribes and ethnic or social groups. Terms for artifacts are not on the list. Similar results, in particular absence of simplex artifact terms, are found by Nichols (2008) for Zuñi.

Domain	Number	Percentage	Example
(i) flora and fauna	13	41.94	*akiwaasas* 'black haw'
(ii) kinship	4	12.90	*-kaa-* 'grandmother'
(iii) body-parts/body-related meanings	2	6.45	*iit-* 'body, corpse, carcass'
(iv) nature-related/topology	3	9.68	*huupirit* 'star'
(v) tribes, ethnic or social groups	6	19.35	*Pasaasi* 'Osage Tribe, Osage Male'
(vi) professions/special persons	1	3.23	*Ctuʔuʔ* 'Witch Woman, a mythological old woman who has supernatural power'
other	2	6.45	*awi* 'fleeting image; quick motion'
total	31		

table 22: Pawnee simplex nouns in the sample according to semantic domain

The analysis for Kiliwa yields quite similar results, summarized in table 23. Here, too, a little less than fifty percent of sampled simplex nouns are flora and fauna terms. In contrast to Pawnee, no kinship terms are found on the sampled pages, but body-parts (somewhat more than in Pawnee), nature-related meanings, and one name of a tribe figure on the list. The noun/verb-ratio in Kiliwa is also quite low, but somewhat larger. Correspondingly, some Kiliwa nouns fall in semantic domains not attested for Pawnee in the sample: there are two native simplex nouns for artifacts, and there is one term denoting an abstract property presumably applicable to many entities.

Domain	Number	Percentage	Example
(i) flora and fauna	14	51.85	*nxil* 'Pitahaya'
(ii) kinship	0	0	-
(iii) body-parts/body-related meanings	6	22.22	*-ha?* 'mouth, voice, breath'
(iv) nature-related/topology	1	3.704	*-kwiy* 'cloud'
(v) tribes, ethnic or social groups	1	3.70	*xwa* 'warrior; enemy, foreigner; principally Cocopa'
(vii) artifacts	2	7.41	*cpat* 'door'
(viii) abstract relations/properties	1	3.70	*cpa?* 'proejction, protrusion, end, tip'
(ix) culture/mode of subsistence/food	2	7.41	*'kuskuwpl* 'edible grass seeds'
total	27		

table 23: Kiliwa simplex nouns in the sample according to semantic domain

Nichols (2008) argues that there are lexico-semantic restrictions in Zuñi as to what a simple noun may denote. In particular, according to her analysis, they are constrained to natural kinds, that is, excluding artifacts. She proposes that this is the explanation for the extremely few loanwords in Zuñi. Given the preliminary results obtained here, this may be true of other North American languages as well, though probably not of all, as Nichols (2008), drawing on data from Brown (1999) also notes that noun borrowability in other Pueblo languages is less constrained (cf. also § 5.4.2.7.1. on borrowing in an American context). Leaving the Americas for Africa to investigate the semantics of simplex nouns in Koyraboro Senni, drastic differences are immediately noticeable. Koyraboro Senni also features many simplex nouns for animals and plants, although the percentage is somewhat depressed when compared with the American data. It has a comparable portion of simplex nouns in the domains of kinship, body-parts, and nature-related terms. From this does not follow that Koyraboro Senni has fewer monomorphemic nouns for animals and plants in absolute numbers, but rather, that their relative percentage is depressed by the presence of monomorphemic terms in other semantic domains. This is noticeable in the domain of artifacts, but particularly obvious in the emergence of terms related to culture, mode of subsistence (the speakers are pastoralists and agriculturalists), and for social relations. Likewise, a term for 'bird' is among the recorded meanings (although it would be wrong to conclude that this is due to increase in societal complexity, since Pawnee also has a simplex noun for 'bird' not on one of the sampled pages). Furthermore, there is a noticeable rise in simplex nouns for abstract relations, properties and quantities. Among

semantic domains of simplex nouns found neither in Pawnee and Kiliwa are those of temporal concepts, such as phases of the day and seasons as well as one noun to denote the emotion 'anger.'[52] Table 25 provides a summary of the Koyraboro Senni data.

Domain	Number	Percentage	Example
(i) flora and fauna	43	36.76	*addihijji* 'aardvark'
(ii) kinship	3	2.56	*feŋge* 'sibling-in-law'
(iii) body-parts/body-related meanings	12	10.26	*diini* 'gums'
(iv) nature-related/topology	8	6.84	*karji* 'thorn, barb'
(v) tribes, ethnic or social groups	4	3.42	*sače* 'ethnic group specializing in leather amulets'
(vi) professions/special persons	1	0.85	*gariibu* 'beggar'
(vii) artifacts	13	11.11	*ferow* 'brick'
(viii) abstract relations/properties	4	3.42	*baka* 'handful'
(ix) culture/mode of subsistence/food	17	14.53	*herow ~ herew* 'young nanny-goat (not yet a mother)'
(x) life-form terms	1	0.85	*subu* 'grass, herb'
(xi) social relations/business	5	4.27	*yaahi* 'friend, pal'
(xii) place names	1	0.85	*bamakoo* 'Bamako'
(xiii) emotions	1	0.85	*zattu* 'desire (for sth.)'
(xiv) temporal concepts	3	2.56	*lahula* 'winter, cold season'
other	1	0.85	*baali* 'pulp (of fruit)'
total	117		

table 24: Koyraboro Senni simplex nouns in the sample according to semantic domain

This trend is continued in Badaga, the language with the highest noun to verb ratio. As table 25 shows, the ratio of flora and fauna terms is further depressed, while the domains of kinship, body-parts and nature-related meanings are relatively constant in their percentages across languages, and the domains of artifacts and abstract relations and properties are represented to about equal percentages in Koyraboro Senni and Badaga. In Badaga, however, there is a dramatic increase in terms having to do with social and religious organization that may be due to an increasingly complex social organization and social stratification. While there are, unlike Koyraboro Senni, no recorded instances of nouns encoding temporal concepts (although they surely must exist), there are many more emotion terms that are encoded nominally rather than verbally in this language, and an additional semantic domain of simplex Badaga nouns not found in the languages discussed so far are units of measurements.

[52] An informal browse through Park and Pratt (2008) reveals that emotions are indeed encoded in Pawnee mostly by verbs, while there are basic nouns for temporal concepts.

Domain	Number	Percentage	Example
(i) flora and fauna	15	11.54	*mundari* 'vine'
(ii) kinship	5	3.85	*auve ~ avve* 'mother, father's wife, wife's father's sister; Toreya term of address for higher-status Badaga women'
(iii) body-parts/body-related meanings	8	6.15	*moḷḷe* 'navel, male nickname'
(iv) nature-related/topology	13	10	*ailu* 'dewdrops, beads of dew'
(v) tribes, ethnic or social groups	2	1.54	*Bekkan* 'Bekkan, Pekkan ...'
(vi) professions/special persons	15	11.54	*haika* 'unintelligent man; male nickname,' 'horseman, equestrian, cavalier; male name'
(vii) artifacts	17	13.08	*moḷe* 'nail, peg, branch'
(viii) abstract relations/properties	4	3.08	*haetu ~ aetu* 'old things'
(ix) culture/mode of subsistence/food	14	10.77	*ha:yi ~ ha:i* 'farmland near a village'
(x) life-form terms	1	0.77	*hakki ~ akki ~ akkilu ~ hakkilu* 'bird, avifauna'
(xi) social relations/business	15	11.54	*saṇḍe* 'war, fight, quarrel'
(xii) place names	2	1.54	*Cocci* 'Cochin ...'
(xiii) emotions	3	2.31	*a:ti* 'wreath; cyclical movement, circular motion, ritual offering'
(xv) units of measurement	2	1.54	*aigua* 'five measures (ca 18.53 litres)'
(xvi) theology	4	3.08	*de:varu* 'god, gods, deity'[53]
other	10	7.69	*saḍunga* 'jingle, jingling sound'
total	130		

table 25: Badaga simplex nouns in the sample according to semantic domain

It has been, intuitively plausibly, claimed that size of vocabulary increases with technological evolution (Witkowski and Burris 1981), and this is congruent given the expansion of specialized cultural vocabulary in Koyraboro Senni and Badaga. However, the methodology Witkowski and Burris employed is dubious: they simply take dictionaries for a number of languages counting the number of entries, and find languages spoken by large industrialized speech communities to have more entries, concluding that "large-scale societies have larger lexicons than small-scale societies" (1981: 144). They acknowledge that dictionary size depends on purpose, but ignore the issue of comprehensiveness and the very different circumstances under which dictionaries for "large" and "small" languages are

[53] This is in fact the plural of *de:va* 'god, godling, deity,' but has its own entry (*de:va* is often used as a honorative singular).

typically created. Specialist vocabulary is said to increase, while 'core' vocabulary remains constant in size. Names for specific plants and animals are said to decrease, and this is consistent with the Badaga results, the one of the investigated languages spoken in the most socially developed speech community.

Summing up, the preliminary evidence from the investigation is that where unanalyzable nouns are few in number, most of them are names of specific animals and plants, with some additional ones in the domains of kinship, nature-related terms, and sometimes artifacts. In languages where they are more frequent, they also cover culture-related meanings (with "culture" perceived in the broadest possible sense), and extend more frequently to also denote abstract concepts as well as emotions. In this context, note that names for animals and kinship terms are precisely the meanings for which unanalyzable basic nouns can be reconstructed for Indo-European, a language in which the nominal lexicon appears to have been characterized by analyzability to a high degree (see § 5.4.2.7.2.).

5.5.2. VERBAL VS. NOMINAL ORIENTATION OF BODY LIQUID AND AEROSOL TERMS

Another aspect of differing lexical organization in terms of nouns and verbs comes from the semantic fields of body liquids and aerosols (as used in physics, i.e. smoke, steam, fog, clouds). In the majority of sampled languages, meanings in both domains are encoded lexically as nouns. However, at times, the morphologically basic expression which encodes them are verbs, not nouns. This is also true of some other meanings in the database. For instance, like a number of other languages in the sample, Ineseño Chumash has a term for 'belt,' *qanatiʹš*, which is derived from the verb *qanatiʹ-* 'to put on a belt.' But differences in the domain of body liquids and aerosols are worth looking at in more detail because they are semantically well-circumscribed, and it is here that differences in lexical organization are most eye-catching. A language in which many attested terms in these domains are basically verbs or derived from other non-nouns is Nuuchahnulth, with the corresponding noun derived from them by the nominalizer *-mis* (which also occurs as a free-standing noun 'thing') or other derivational suffixes:

(10.) a. 'cloud': *ɬiẇaḥmis* /ɬiẇaḥak-mis/ 'be.cloudy-NMLZ'
b. 'fog': *ʔučqmis* /ʔučqak-mis/ 'foggy-NMLZ'
c. 'smoke': *qʷiš-aa* 'to.smoke-??'
d. 'steam': *muqckʷii* /muq-ckʷɨ̌/ 'to.steam-remains.of'
e. 'blood': *ḥis* 'blood, to bleed;' *ḥis-mis* 'blood/bleed-NMLZ'
f. 'saliva': *taaxckʷi* /taaxʷ-ckʷɨ̌/ 'spit-remains.of'
g. 'sweat': *ƛ̓upẏiiḥa-ckʷim* 'to.sweat-??'
h. 'snot': *ʃintmis* 'snot, nasal mucous'

The only meaning not (also) encoded as a basic verb is that for 'snot,' though note that it, too, ends in *-mis*, although there is no corresponding verb *ʃint* in the consulted source. Note that in Nuuchahnulth the root *ḥis*, which bears the semantic content of 'blood,' is ambiguous as to its lexical category and can function as both noun and verb, and that there exists an overtly nominalized version of this which singles out the referential read-

ing. The Nuuchahnulth source does not contain counterparts for the meanings ‘pus’ and ‘urine;’ verb-based terms for these meanings are found for instance in Chickasaw (*kalha-'* ‘have.pus.come.out-NMLZ’) and Sora (*'aŋ(ɲ)um-ən* ‘urinate-N.SFX’).

Percentages for terms like those in (10.) for all languages are in Appendix C (note that only plain, semantically inert derivation serving only to change the lexical category is counted here; thus Pipil *te:mal* ‘pus,’ which is derived from *te:ma-* ‘to fill,’ and similar terms are not counted, as are semianalyzable terms of all kinds).

The map in figure 38 shows the distribution and strength of the phenomenon visually. As the map shows, such terms are relatively frequent in North America, which can also be observed in the boxplot in figure 39. However, statistically, the areal differences are not quite significant under the Dryer-6 breakdown ($\chi^2 = 10.3366$, df = 5, $p = .06624$, Kruskal-Wallis rank sum test).

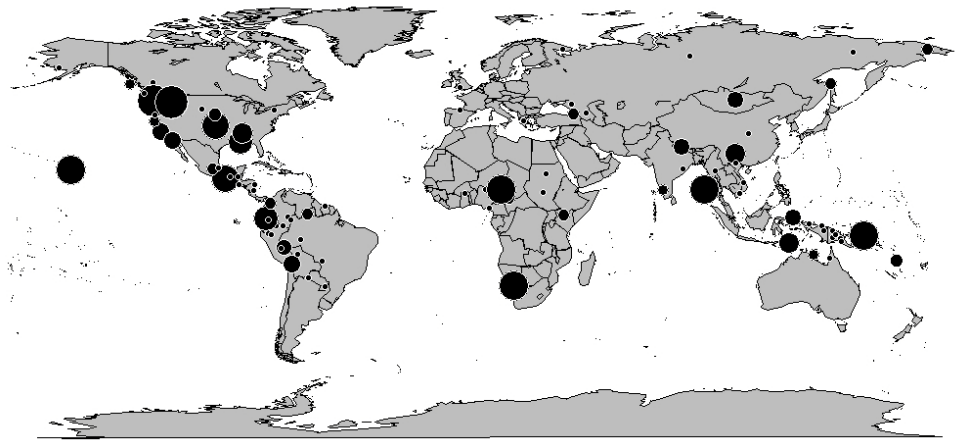


fig 38: the distribution of deverbal or N/V-ambiguous terms for aerosols and body liquids, core sample

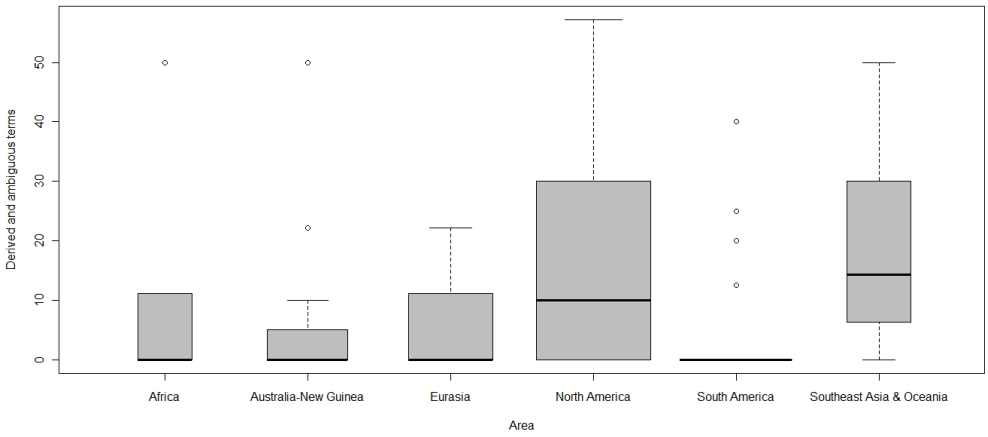


fig. 39: Percentage of deverbal or N/V-ambigous terms in different areas, using Dryer's (1992) breakdown

The picture is quite similar when only overtly marked deverbal terms are considered. The corresponding map is seen in figure 40.

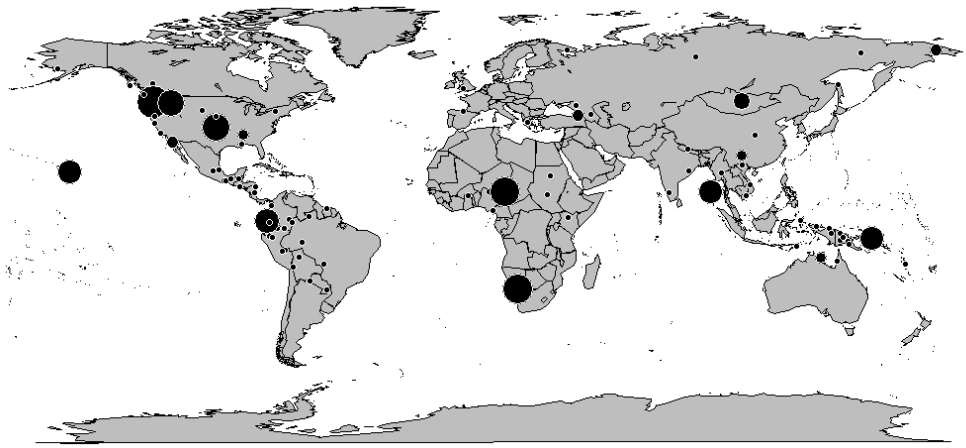


fig. 40: the distribution of deverbal terms for aerosols and body liquids, core sample

In spite of the fact that here North America stands out even more clearly when it comes to verb-based lexical categorization of the relevant meanings, as also seen in the boxplot in figure 41, areal biases are not significant statistically (χ^2 = 3.4827, df = 5, p = .626, Kruskal-Wallis rank sum test).

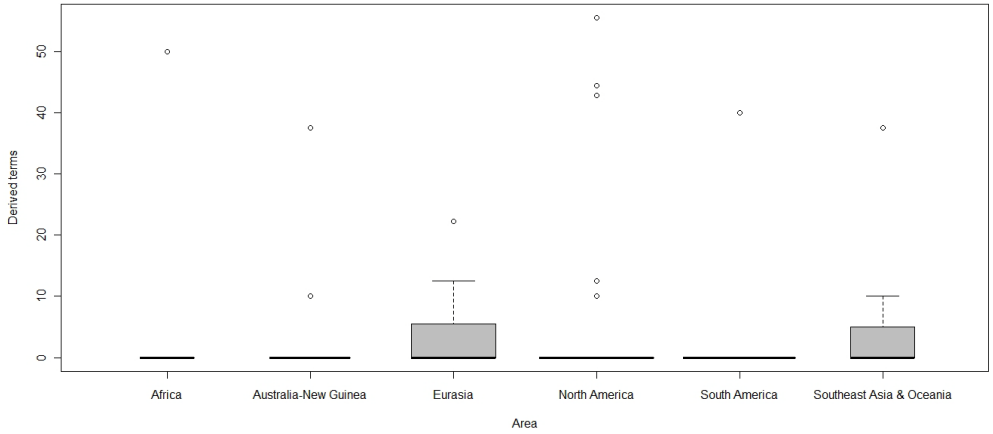


fig. 41: Percentage of deverbal terms in different areas, using Dryer's (1992) breakdown

Unsurprisingly, there is a strong correlation in both cases with the percentage of derived terms for the other meanings considered (Spearman's ρ = 3086653, p = .005968 when taking into account both derived and ambiguous terms and Spearman's ρ = .3461545 and p = .001907 when only taking into account derived terms).

5.6. A NOTE ON COLEXIFICATION, ANALYZABILITY, AND PHONOLOGY

Could it be possible that the phonological structure also is responsible for the degree of polysemy or at least for certain patterns of colexification, such as 'river' - 'water'? This is suggested both by the case study of Vanimo as well as the situation in Northwest Caucasian, where it has been noted that simplex lexical items have a rather broad denotational range to compensate for the limited number of roots the phonological system enforces. Thus it is a conceivable situation that a high degree of analyzability goes hand in hand with a high degree of simplex lexical items with broad reference, in other words, lexical items that colexify meanings that would be expressed by morphologically unrelated lexical items in other languages. As noted in chapter 3, next to more general issues having to do with the extraction of colexification, there is an effect of the type of the consulted source on the quantitative measure of colexification so that testing on the entire statistics sample is not feasible. However, it is possible to narrow down the sample even further by removing data from languages for which the source is of the kind that influences this percentage statistically. However, even with this measure taken, statistically no interaction of the percentage of colexification was found with any of the phonological features under scrutiny. In contrast, there is a correlation between the degree of analyzability and the degree of colexification (values for both are in Appendix B) for those languages where there is no statistical bias on the measured degree of colexification due to the nature of the consulted source. This analysis shows that, on average, languages with a high degree of colexifying lexical items also tend to have low degrees of analyzability, while languages

with a comparably low percentage score when it comes to colexification, typically have a more analyzable lexicon (p = .0018 by a Mixed Model design). Thus, rather than a upward trend in the degree of colexification that is correlated with a rise in the number of analyzable lexical items in the investigated vocabulary, there is an inverse relationship between colexification and analyzability: The more analyzability, the less colexification and vice versa. The correlation is plotted in figure 42.

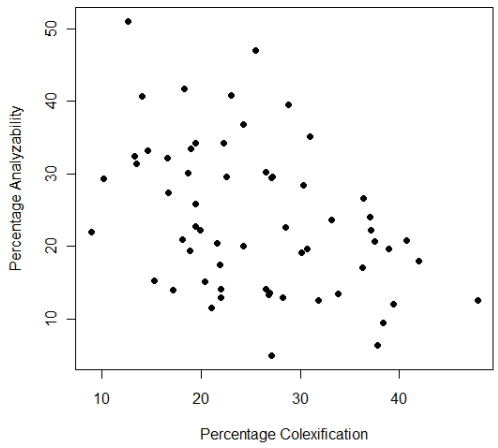


fig. 42: Correlation between the measured percentage of colexification and analyzability

This mirrors the basic observation from § 3.5.: the same semantic relationship may be expressed by colexification in some languages and by analyzable lexical items in others.

5.7. METAPHOR AND METONYMY

5.7.1. INTRODUCTION

Rather than looking at the quantitative aspect of lexical motivation, with which most of the discussion in this chapter has been concerned so far, this final section looks at the semantic side of things, in particular contrasting the degrees to which languages employ metaphor or metonymy as defined in chapter 3 as semantic relations. For quantitative evaluation, these differences are measured by the CONTIGUITY-SIMILARITY RATIO, which is calculated by dividing the relative percentage of lexical items motivated by similarity by the relative percentage of lexical items motivated by contiguity. Hence, a value of 1 indicates that the two values are in balance, a value smaller than one indicates that contiguity dominates (the smaller the value, the stronger this dominance is) and a value larger than one that similarity is the dominant semantic relation in a given language (again, the larger the value, the stronger the dominance). Values for this ratio are in Appendix B. The map in figure 43 plots the cross-linguistic differences in this area for the languages of the core sample.

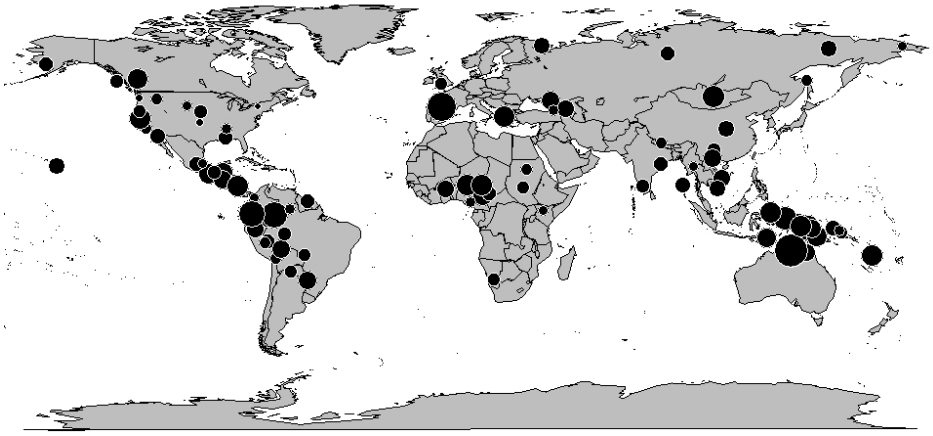
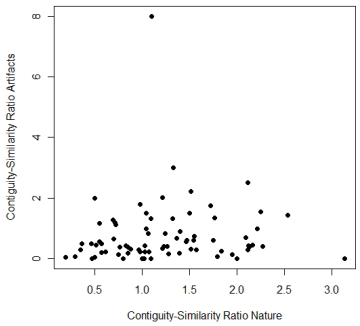


fig. 43: differential degree of metaphor- vs. metonymy in motivated terms, core sample

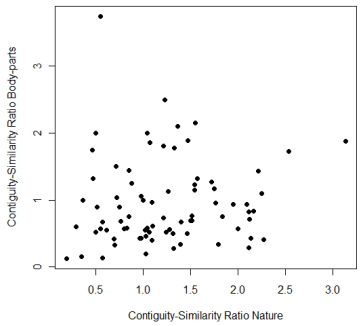
A question raised by Koch and Marzo (2007: 273) is: "Are there predominantly metaphorical languages?," in other words, whether there is any non-random signal in the distribution of the variable as seen in figure 43. The answer to this question is, as the following discussion will show, yes, there seem to be, but the much more interesting question to ask is, why?

5.7.2. CORRELATIONS BETWEEN THE PROFILE OF LANGUAGES IN DIFFERENT SEMANTIC DOMAINS

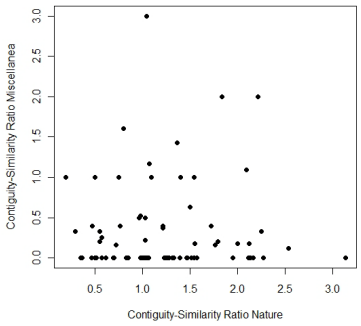
The plots in figure 44 visualize differences of metaphorical vs. metonymic semantic relations in the languages of the statistics sample across semantic domains (see Appendix B for data), and tests for each possible combination for correlations between the domains, with the Spearman's ρ being approximate due to ties and *p*-values corrected using Bonferroni corrections as implemented in R due to multiple testing.



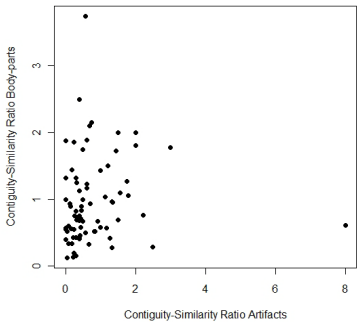
Nature vs. Artifacts: ρ ≈ .18, *p* ≈ .60



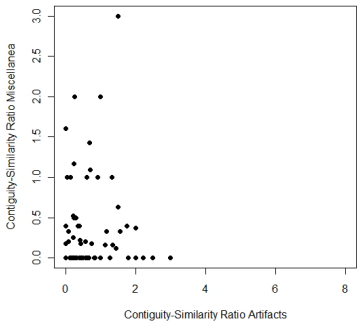
Nature vs. Bodyparts: ρ ≈ .14, p ≈ .94



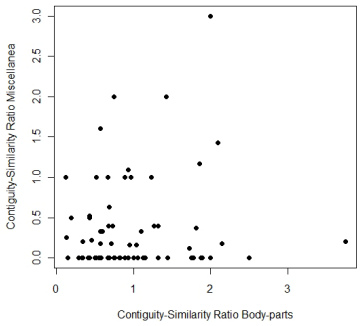
Nature vs. Miscellanea: ρ ≈ -.02, p = 1



Artifacts vs. Bodyparts: ρ ≈ .29, p ≈ .07



Artifacts vs. Miscellanea: ρ ≈ -.03, p = 1



Bodyparts vs. Miscellanea: ρ ≈ .09, p = 1

fig. 44: the contiguity-similarity ratio across semantic domains

What this analysis shows is that under all other possible pairings, the correlation is not significant, thus meaning that here the distribution is less clearly paired when comparing it with the results for the degree of analyzability reported in § 5.2.2.

However, next to asking about domains and the differential degree to which contiguity- and similarity-based denominations are found, it is also possible to ask whether there is any difference with respect to their subtypes as established in § 3.6.2.2. The boxplot in fig. 45[54] shows that the ratios of terms where the relation of functional similarity as opposed to perceptual similarity, is found to the highest degree is that of artifacts (values are in Appendix D).

[54] The extremely high value of 276.5 for the meaning 'bark' lies outside the plotted area.

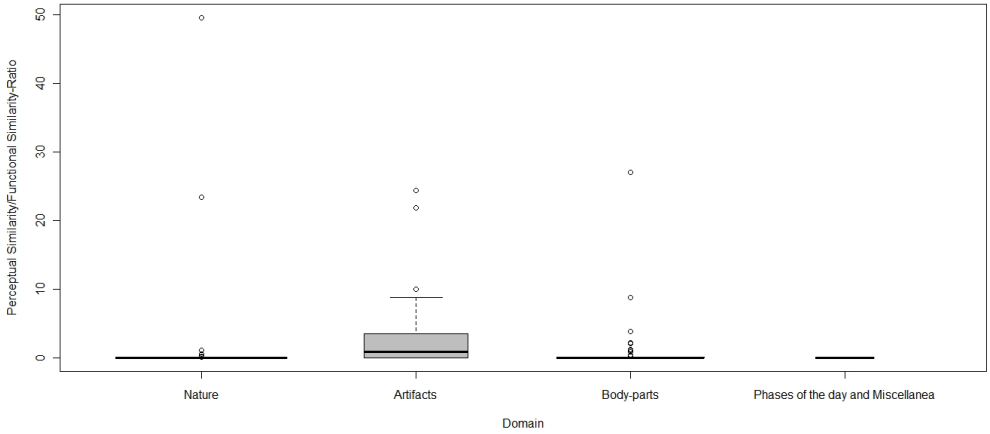


fig. 45: differences across semantic domains in functionally- vs. perceptually-based similarity

This is mostly due to the distribution of the relation of functional similarity within semantic subdomains. Compare, for instance, as examples of morphologically complex terms, Hausa *jurgi-n sama* 'boat/train-GEN sky' = 'airplane' as well as Kaluli *ho:n ko:su* 'water airplane' = 'power boat, boat introduced during colonial contact.' In the domain of tools, frequently colexification of meanings that have functionally similar referents are found, for instance Jarawara *yimawa* 'knife, machete' (cf. also Sko *tàng*, glossed as 'sickle, knife, machete, general term for blade of any kind' and note in this regard that the distinction between genuine polysemy and semantic generality is not at stake presently) and Sentani *o'bi* 'ladder, stairs.' Another frequent pattern is colexification of 'house' and 'nest' (in spite of the equally if not more common pattern for 'nest' to be named by a morphologically complex term 'bird house,' see Appendix E, 41), and abstract extension of meanings such as that of 'street' or 'way' to 'method' or 'manner' (see Appendix E, 92).

A similar result, with artifact-terms standing out, is obtained when one does not look at the difference between the two different types of similarity-based relations, but instead compares for each concept whether contiguity-based or similarity-based conceptualizations, as measured by the contiguity-similarity ratio, abound. As the plot in figure 46[55] shows, it is again the domain of artifacts in which particularly low values for the contiguity-similarity ratio, that is, prevalence of contiguity as the semantic relation is found.

[55] Again, the meaning 'bark' is with a value of 111 outside the plotted area.

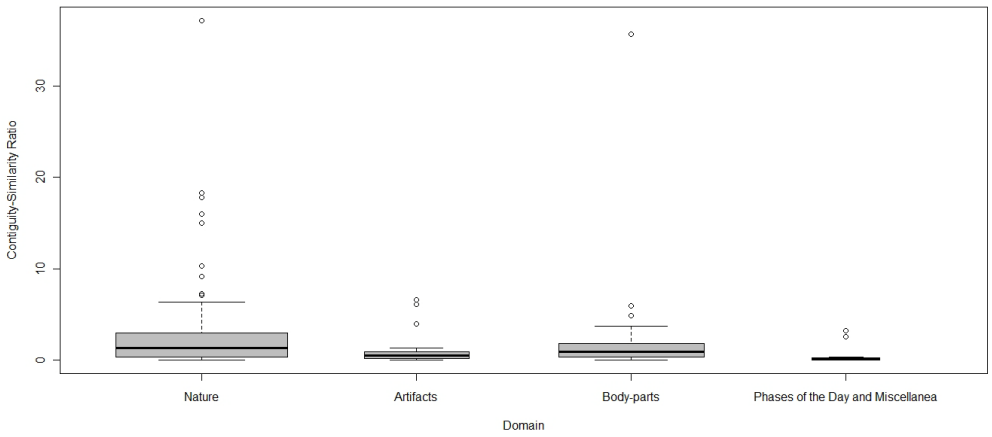


fig. 46: differences across semantic domains in prevalence of contiguity and similarity-driven conceptualizations

However, the miscellanea-domain shows a similar behavior, which is in all likelihood due to contiguity-based associations for meanings such as 'day' and 'night.' For instance, 'day' is, by contiguity, frequently colexified with 'sun,' and 'night' with 'dark.' Moreover, contiguity-based terms for 'noon' such as Kildin Saami *piejjv-kēssk* 'day-middle' abound (see Appendix E, 151, 153, and 154 for fuller discussion).

5.7.3. INFLUENCES OF STRUCTURAL FACTORS?

Analogously to the data on the degree of analyzability and the type of analyzable lexical item, preliminary tests were carried out on the basis of the data in the World Atlas of Languages Structures to elucidate possible interactions between structural features and the dominance of either contiguity or similarity as the semantic relation underlying analyzable items and colexification. Significant *p*-values (all by Kruskal-Wallis Rank sum tests, and again, given the exploratory nature of the tests, uncorrected for multiple hypothesis testing) were obtained for the following features:

(i) Voicing in Plosives and Fricatives (Maddieson 2005g):
$\chi^2 = 10.1558$, df = 3, $p = .01729$
(ii) Uvular Consonants (Maddieson 2005f):
$\chi^2 = 7.2236$, df = 2, $p = .02700$
(iii) Lateral Consonants (Maddieson 2005c):
$\chi^2 = 11.9302$, df = 4, $p = .01788$
(iv) Politeness Distinctions in Pronouns (Helmbrecht 2005):
$\chi^2 = 3.7231$, df = 1, $p = 0.05367$
(v) The past tense (Dahl and Velupillai 2005):
$\chi^2 = 8.6491$, df = 2, $p = 0.01324$

(vi) Order of relative clause and noun (Dryer 2005e):
χ^2= 8.1259, df = 3, p = 0.04348
(vii) Position of Interrogative Phrases in Content Questions (Dryer 2005h): χ^2= 5.4938, df = 2, p = .06413
(viii) Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun (Dryer 2005k):
χ^2= 9.2295, df = 3, p = 0.02639
(ix) Alignment of case marking of pronouns (Comrie 2005):
χ^2= 10.5261, df = 5, p = 0.06163
(x) Zero Copula for predicate nominals (Stassen 2005b)
χ^2= 3.7386, df = 1, p = .05317
(xi) Tea (Dahl 2005): χ^2 = 5.7909, df = 2, p = 0.05527

Of these, all but three features remained significant under a Mixed Model design controlling for areal effects. The remaining eight features are:

(i) Voicing in Plosives and Fricatives (Maddieson 2005g) : p = .0209
(ii) Uvular Consonants (Maddieson 2005f): p = .0172
(iii) Lateral Consonants (Maddieson 2005c): p = .0263
(iv) Politeness Distinctions in Pronouns (Helmbrecht 2005): p = .0378
(v) The past tense (Dahl and Velupillai 2005): p = .0029
(vi) Order of relative clause and noun (Dryer 2005e): p = .021
(vii) Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun (Dryer 2005k): p = .0067
(viii) Alignment of case marking on pronouns (Comrie 2005): p = .0113

Cross-validating the results on the basis of the validation sample was not possible for features (i), (ii), (iii), (v), and (viii).[56]

There was a replicable difference between languages with no and a binary politeness distinction in pronouns (estimate of the validation sample 0.46 as opposed to 0.4503 ± 0.1862 in the original sample), which is plotted in fig. 47.

[56] The available estimates for the sake of completeness are: (i): -.1650 vs. .346 ± .1236, .08 vs. .3627 ± .2141 and .06 vs. .6293 ± .2141; (ii): -.2533 vs. -.40640 ± .14934; (iii): .5333 vs. .27476 ± .14366 and .185 vs. -.14190 ± .22714; (v): .02 vs. -.678 ± .1442; (viii): .0660 vs. .395 ± .1225.

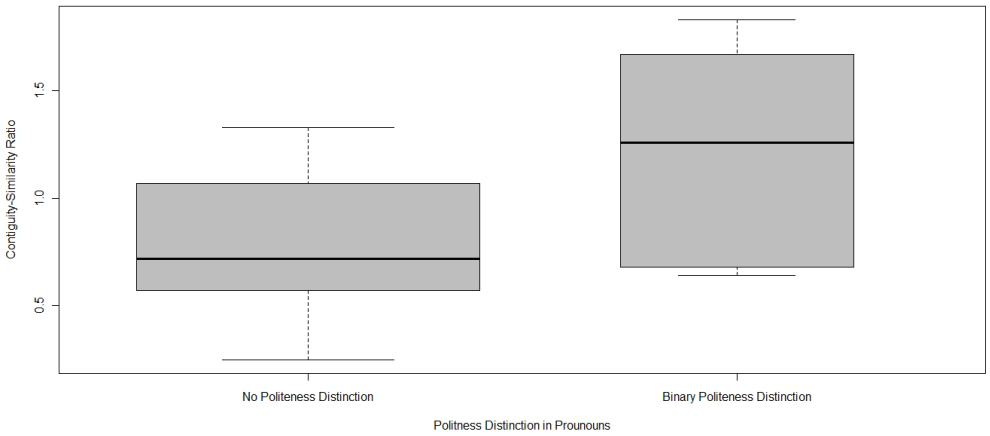


fig. 47: differences in the contiguity-similarity ratio depending on politeness distinctions in pronouns

Moreover, there are two replicable correlations which have to do with the order of relative clause and noun. As the associated plots in figures 48 and 49 show among other information, metaphor-based associations are more common in languages in which relative clauses precede the noun.

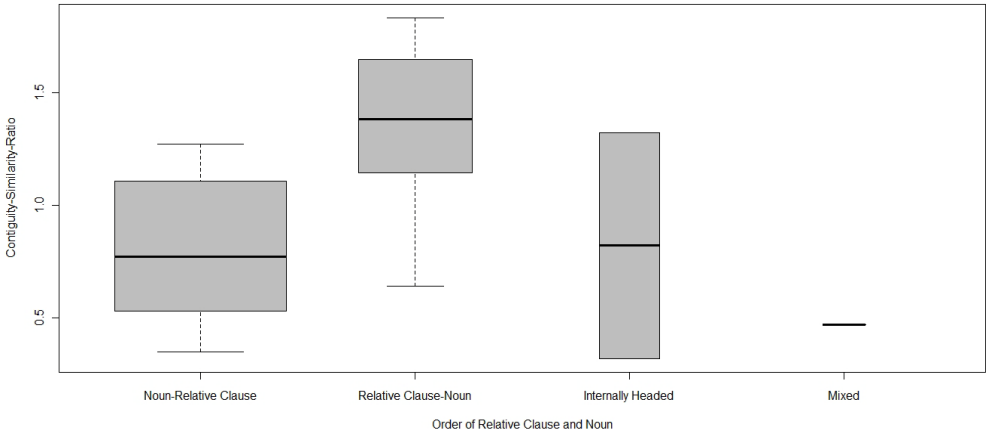


fig. 48: differences in the contiguity-similarity ratio depending on the order of relative clause and noun

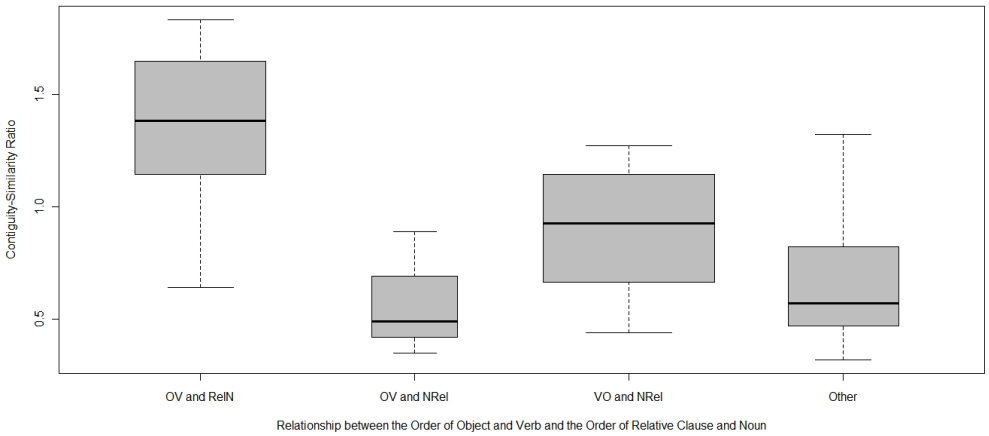


fig. 49: differences in the contiguity-similarity ratio depending on the order of object and verb and the order of relative clause and noun

Generally, these correlations are difficult to make sense of, and hence they are for the time being simply mentioned without an attempt at an explanation. It needs to be borne in mind that, as already noted in the discussion above, the overlap between the WALS samples and the present sample is at times rather small, and hence so is the empirical datapool from which generalization may be drawn, to the effect that the behavior of few individual languages can lead to the emergence of statistical significance. Conversely, also because of these facts, some genuine interaction in fact may exist for features for which none has been diagnosed. At any rate, with politeness distinctions in pronouns and the order of relative clause and noun as the only candidates, the influence of structural features coded in WALS on semantic relations underlying motivated items in the lexicon appears to be small.

5.7.4. NO STRONG AREAL EFFECTS ON THE RELATIVE DEGREE OF METAPHOR AND METONYMY

Areal effects are not very pronounced either, and where they exist, there are relatively straightforward explanations. Figure 50 plots the results of the relative degree of metaphorical expressions using Dryer's 6-way breakdown of the world. This is for the time being simply for the purpose of illustration; the difference is not significant statistically ($\chi^2 = 5.2236$, df = 5, $p = .3892$, Kruskal-Wallis rank sum test).

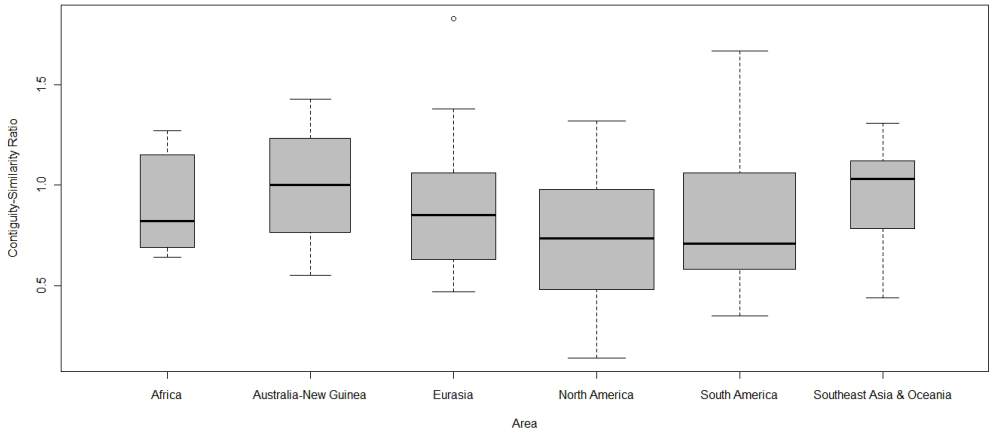


fig. 50: Areal breakdown of the relative degree of metaphor-driven semantic relations, using Dryer's (1992) breakdown.

The lowest degree of metaphor-based conceptualization is found in the Americas, where contiguity as a semantic mechanism in colexification and analyzable lexical items prevails, although it is not significantly more dominant here than elsewhere. There was also no evidence for areal differences under the other two standard breakdowns used in the present study (Nichols-11: $\chi^2 = 11.9601$, df = 10, $p = .2877$; Nichols-3: $\chi^2 = 3.8834$, df = 2, $p = .1435$, both by Kruskal-Wallis rank sum tests).

Testing for individual semantic domains yields almost always negative results under all testing conditions, with the exception of artifacts, which have a significantly different relative degree of metaphor and metonymy at $p = .02277$ ($\chi^2 = 7.5644$, df = 2, Kruskal-Wallis rank sum test) under the Nichols-3 breakdown, plotted in figure 51.

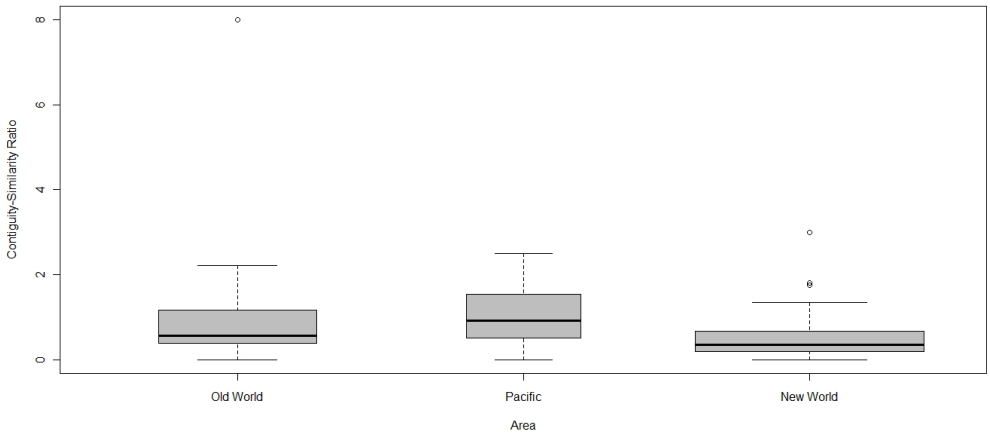


fig. 51: Areal breakdown of the relative degree of metaphor-driven semantic relations in artifacts, using Nichols's (1992: 27) breakdown

This plot essentially replicates the one when all semantic domains are considered: high degree of metaphorical relations in the Pacific area, with depressed ratios in the New World. This is most likely due to two factors: first, the languages of the Americas have a high ratio of motivated, in particular analyzable terms for artifacts, and these tend to be named with reference to their function, which is by definition a relationship of contiguity, not one of similarity. Perhaps more importantly, as will be seen in the following section, there is an overall correlation between the preference for terms of the derived rather than of the lexical kind to be driven by contiguity, and it is again in the Americas where languages of this type cluster.

This negative outcome should not be too surprising, if one bears in mind the ultimate cause of areal effects: the need of bi- or multilingual speakers to increase intertranslatability between the languages they speak (e.g. Gumperz and Wilson 1971), and the need to express the same thought in two different languages (Sasse 1985). As noted by both Gumperz and Wilson and Sasse, this single need underlies contact phenomena in morphology and syntax, but are equally responsible for convergence in semantics and lexicon. Since relative degree of metaphor is a highly abstract measure, it seems unlikely to be influenced by areal factors as it is not directly manipulateable by speakers. Rather, contact effects are clearly recognizable in the denominations of individual meanings and their semantic structure (see § 6.4.3).

5.7.5. METAPHOR AND METONYMY AND PREFERRED TYPE OF ANALYZABLE LEXICAL ITEM

If the degree of metaphor and contiguity does not appear to be decisively influenced by grammatical factors nor for the most part by areal factors, is their distribution completely random? In fact, there appears to be a structural factor that triggers the languages' behavior in this regard.

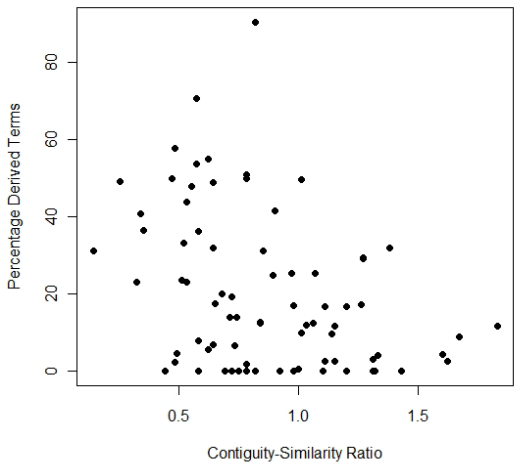


fig. 52: Correlation between the contiguity-similarity ratio and the percentage of derived terms of all analyzable terms

There is a non-trivial and significant effect of the relative frequency of complex lexemes of the derived and of the lexical type as defined in § 3.6.1. and the fundamental semantic relation -similarity or contiguity- that dominates the lexicon (*p*-value associated with predictor = .0004 by a Mixed Model design controlling for areal affects, estimate = -.0066). An illustrating plot is in figure 52.

Why is that? As mentioned in § 4.4.2., the Central Yup'ik postbase *-yak* 'thing similar to' is by its semantics prone to create complex lexical items where the referent of the complex terms stands in a relation of similarity to that of the derivation base, so it is logically perfectly possible to have similarity-based terms derivatives. Some examples from another language, Muna, which is one of the rare languages that have several terms of this kind distributed over all semantic domains, are in (11.):

(11.) Similarity-based derivatives in Muna
a. *ka-mbea* 'ABSTR-shine' = 'flower'
b. *ka-ofe ~ ka-ufe* 'ABSTR-squeeze.rice.in.round.shape' = 'nest'
c. *kara-kara* 'yard.fence-RED' = 'rib'

However, derived terms in most languages are not metaphorical in nature, but have a metonymic basis (see also Anderson 2011b: 285). This lies in the very nature of the process, more precisely, in the semantics of derivational morphemes found in many languages that often serve to derive names for instruments or locations from the derivation base (see Bauer 2002 for a cross-linguistic survey of the semantics of derivational morphemes, which includes a number of more unusual meanings, but none that is susceptible to establish a relation of similarity with the meaning of the derivation base in particular). Furthermore, derivatives typically do not allow for contiguity anchoring, leading to a "cognitive leap" that appears to be dispreferred cross-linguistically. Nuuchahnulth derivatives may serve as examples for the overwhelmingly contiguity-based derivatives in the world's languages:

(12.) Contiguity-based derivatives in Nuuchahnulth
a. *maamaati* /maa-mat-ǐʻp/ 'RED-fly-THING...ED' = 'bird'
b. *hiɬ-waḥsuɬ* 'LOC-go.out.from' = 'estuary'
c. *ƛupkẏak* /ƛupk-ẏak^{w}/ 'untie-INSTR' = 'key'

And even in Muna, the locative nominalizer *ka-* occurs typically in contiguity-establishing function, such as in *ka-bhawo* 'mountain' (*bhawo*, 'high').

However, the correlations are found for semantic relations in the lexicon as a whole, that is including those in morphologically complex lexemes as well as those in colexification. Since the above observations pertain exclusively to analyzable items, it is necessary for the present purposes to distinguish between semantic relations in analyzable items and those in colexification and to assess their behavior separately. When this is done the picture becomes much clearer. Then, there is a highly significant correlation at p = .0001 between the dominant type of complex lexical item (derived vs. lexical) and the predilection for similarity-driven as opposed to contiguity-driven semantic relations in analyzable terms (data are in Appendix B) under the same Mixed Model design controlling for area, with the value for the contiguity-similarity ratio logarithmically transformed. Figure 53 illustrates the correlation.

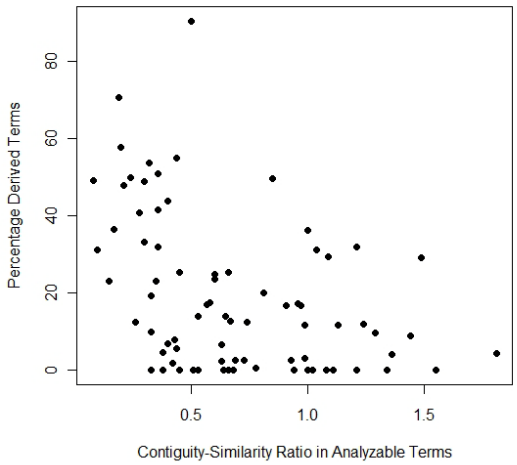


fig. 53: Correlation between contiguity-similarity ratio and the percentage of analyzable terms of the derived type

But what about colexification? It would be a spectacular finding if the same preference as in analyzable lexical items would extend to colexification as well. However, this is not so. There is no effect of the dominant type of complex lexical items on the semantic relation in colexification (p = .8858) under a Mixed Model design with the contiguity-similarity ratio in colexifying terms (values are again in Appendix B) logarithmically transformed, showing that the overall correlation between the variables is entirely due to the semantic relations in analyzable terms, with the relations due to colexification in fact confounding the picture.

It is possible to align this finding with the previous discussion, since both lack of verbal person marking and lack of an elaborate derivational apparatus are characteristic of a language type known as "isolating" in traditional morphological typology. Such languages, because of their restricted bound morphology, will to a great degree make use of lexical rather than derivational resources to coin their morphologically complex expres-

sions. In this sense, the observed patterns point to the TENDENCY OF ISOLATING LANGUAGES TO MAKE USE OF METAPHOR AS A LEXEME-INTERNAL SEMANTIC RELATION IN ANALYZABLE TERMS TO A GREATER DEGREE THAN NON-ISOLATING ONES.

The differences can be exemplified by contrasting data from Austro-Asiatic languages. This family consists of two major branches, the Munda languages spoken on the Indian subcontinent, and the Mon-Khmer languages, most of which are spoken in Southeast Asia. The split between these two primary branches is deep, and consequently, Munda and Mon-Khmer languages are also quite different typologically. Munda languages have rich verbal morphology; in contrast, Mon-Khmer languages participate in the Southeast Asian Sprachbund and exhibit its typical features: they are tonal and are largely isolating, thus for instance not featuring person agreement on the verb. These typological differences are mirrored in the lexicon, and they can be shown by contrasting data from Sora (Munda) and Sedang (Mon-Khmer). In Sora, about 25 per cent of analyzable terms in the data are of the derived type. Examples include:

(13.) a. *'ge:mən* /ge:m-ən/ 'to.light-N.SFX' = 'flame'
b. *gərob'go:b-* /g<ər>ob-gob-/ '<INSTR>sit-RED-'= 'seat'
c. *meme:-n* 'suck-N.SFX' = 'breast'
d. *gag'garən ~ gal'galən /gag-gar-ən/* 'RED-pierce/bore.a.hole-N.SFX' = 'scar'

(Of course, Sora also features complex expressions that are not based on verbs, such as *'kuru:-tam-ən* 'body.hair-mouth-N.SFX' for 'beard'). In the lexicon in general as in the examples in (13.), contiguity-driven conceptualizations outnumber metaphor-driven ones, as indicated by the contiguity-similarity ratio of .79 (.4 in analyzable terms only).

In line with the observations made above, there is a correlation between dominance of complex expressions of the lexical type and metaphor as the conceptual mechanism underlying lexical motivation. In the Mon-Khmer language Sedang, one encounters roughly the reverse situation. The percentage of derived terms is, with 11.9 per cent of all analyzable terms, only about half of that encountered in Sora. In absolute figures, this amounts to a number of only two deverbal terms in the data available for Sedang, formed using the nominalizing infix <ơn>, for instance *kơnep* 'scissors' (*kep*, 'to cut hair'). Examples of analyzable terms of the lexical type in Sedang are in (14.).

(14.) a. *kia hia* 'ghost light.weight' = 'clouds, air, smoke'
b. *kơtôu ma* 'bark/rind/shell eye' = 'eyelid'
c. *tróang mơhéam* 'road blood' = 'blood vessel, vein, artery'
d. *tea ma* 'water/liquid eye' = 'tear'

The chosen examples are roughly representative of the relative degree of contiguity and similarity as underlying processes: while there clearly are complex terms that are contiguity-driven, such as (14d.), similarity-based complex lexemes outnumber them in the vocabulary segment under investigation, as indicated by the contiguity-similarity ratio of 1.03 (1.24 in analyzable terms only).

5.7.6. A FURTHER POSSIBLE FACTOR

There is a strand of recent research in Social Psychology that may turn out to open up extremely interesting prospects for a better understanding of preferences between languages for the prevalence of semantic associations they favor. Cognitive Psychologists distinguish two types of reasoning, one is based on intuitions on the basis of gathered experience and is associative in nature, the other is categorical, logical and operates by the application of rules (Sloman 1996). The former system is based on relations of spatiotemporal contiguity and similarity, the latter on categorical and taxonomic relations. Importantly, although both systems are probably available to all humans, there are marked cross-cultural differences in the prevalence of each of the systems in reasoning. In particular, the former, associative system is dominant in Asia, while in languages of Western cultures, the taxonomically oriented system is employed with greater frequency (Norenzayan et al. 2002, Nisbett 2003, see Norenzayan et al. 2007: 577-586 for review). For instance, Masuda and Nisbett (2001) show that Japanese subjects remember more of the background of an artificial underwater scene they were shown, and started descriptions of the scene by introducing the background, Westerners were more likely to separate a particularly salient target object - a "focal fish," which is bigger and more colorful than other elements - of the scene. Ji et al. (2004) show that prevalence of one of the two systems has effects in linguistic tasks specifically: in a triad categorization task, American subjects were more likely to group sets of words together on the basis of category structure (for instance, grouping 'monkey' together with 'panda'), while Chinese-speaking subjects were more likely to group referents together on the basis of them sharing the same frame (e.g. 'monkey' and 'bananas'). The Chinese subjects had some degree of proficiency in English, and were tested using both English and Chinese; the effects remained noticeable regardless of this difference.

Thus, given the areal distribution of the dominance of the systems, it may be the case that in languages of Western cultures, motivated terms, in particular neologisms, may be characterized by reflecting taxonomic structures, as e.g. in endocentric compounds, while denominations in Asian languages could be expected to be of an associative, contiguity and/or similarity-based (as e.g. in exocentric compounds). In the areal breakdown in figure 50, one can observe that there is a higher number of metaphor-driven lexical associations than in languages of Eurasia and Europe. However, it is not entirely clear whether the distinction of contiguity vs. similarity as presently defined is in fact the adequate measure to bring to light such putative influences in language, since the associative system operates both on the basis of spatiotemporal contiguity and family resemblances (metaphor), and it might be more profitable indeed to approach the question distinguishing between e.g. endocentric and exocentric compounds.

Moreover, the question of whether there are indeed cross-linguistic effects of the two types of reasoning on lexical structure unfortunately cannot at this point of time be elucidated in more detail because "[l]ittle is known about the operation of these two systems of reasoning across diverse cultural groups" (Norenzayan et al. 2002: 654), in spite of some evidence that, rather than a difference between Western and East Asian societies, on contrasting which research has focussed so far, the difference really is between the industrialized West and the rest of the world as well as differences based on mode of subsis-

tence (Henrich et al. 2010). However, these differences in cognitive styles (to take up a term by Hymes 1961) and their effects on linguistic tasks demonstrated by Ji et al. (2004), which are in turn likely based on different patterns of social, political, and personal organization (Nisbett et al. 2001), suggest that it is possible that there are cultural effects on the structuring of the lexicon that would provide evidence against the claim uttered for instance by Alinei (2001) that languages randomly pick features of referents in naming them and that goes beyond the trivial sense of contingent aspects of material culture, as when, say, a language colexifies 'thorn' and 'needle' because thorns are used as needles.

5.7.7. SUMMARY

Given that influences of cognitive reasoning cannot be systematically checked at the present state of knowledge, the overall conclusion for the time being thus is that the DOMINANT WORD-FORMATION DEVICE INFLUENCES WHETHER THE LANGUAGE WILL FAVOR CONTIGUITY- OR SIMILARITY-BASED DENOMINATIONS IN MORPHOOGICALLY COMPLEX LEXICAL ITEMS. This is a non-trivial finding, since, to reiterate, there is no a priori reason that compounds must be metaphorical and derivatives must be metonymic semantically. It is also a highly interesting finding because, put in other words, one can observe here that languages, depending on the nature of aspects of their grammar (i.e. word-formation), carve up the essentially same or near-same reality, as represented by the meanings on the list which are presently studied, in quite different ways. At any rate, it would be highly interesting to expand the findings empirically in concrete fieldwork to ascertain the soundness of the semantic side of the analysis. As pointed out by Aikhenvald (2007: 9), "compounding is widespread in isolating languages, while derivation is a property of languages of other types; this follows from the tendency to have a one-to-one correspondence between a morpheme and a word in isolating languages." It is therefore no coincidence that high rates of contiguity-based semantic relations at the expense of similarity-driven ones are dominant in the Americas, because here derived-type languages concentrate (though note that the correlation is not due to this fact alone, since area is controlled for). The typology can now be enhanced and finalized in table 26 by adding a lexico-semantic correlate to the lexical and derived types: that of predilections for similarity-based and contiguity-based semantic relations in morphologically complex lexical items respectively.

	High degree of Analyzable Terms	Low Degree of Analyzable Terms
Lexical Dominating, Derived Subsidiary	• Low complexity in verbal person marking, fixed word order	• Low complexity in verbal person marking, fixed word order
	•Simple phonology, short roots	•Complex phonology, long roots
	•Dominance of similarity as a semantic relation in analyzable terms	•Dominance of similarity as a semantic relation in analyzable terms
	•Tentatively: favors neologisms	•Tentatively: favors borrowing
Derived Dominating, Lexical Subsidiary	• High complexity in verbal person marking, free word order	• High complexity in verbal person marking, free word order
	•Simple phonology, short roots	•Complex phonology, long roots
	•Dominance of contiguity as a semantic relation in analyzable terms	•Dominance of contiguity as a semantic relation in analyzable terms
	•Tentatively: favors neologisms	•Tentatively: favors borrowing

table 26: final cross-classification of language types summarizing the established correlations

With this table, the quantitative evaluation comes to an end. It is summarized in textual form in the final section that is to follow.

5.8. CHAPTER SUMMARY

This chapter presented a quantitative evaluation of the variables surveyed in this work, and tried to establish correlations with language-internal structural as well as some social and cognitive factors and to provide, where they are found, an explanation for the observations. It turned out that most of the obviously relevant factors that interact with the degree of analyzability of the nominal lexicon is structural rather than areal-typological (borrowing etc.), sociolinguistic (L2 learners) or cultural (word taboo, mode of subsistence). More precisely, the structural factors involved are mostly phonological: the simpler the syllable structure, the smaller the consonant inventory, the shorter the monomorphemic native lexical morpheme, the more analyzable terms the sample languages have in their nominal lexicon. Another relevant phonological factor is suprasegmental: tone, such that tonal languages are likely to be characterized by a higher degree of analyzability in the lexicon that non-tonal ones. Each of the factors alone was found to be significant, but due to cross-linguistic dependencies between themselves, it is not entirely clear whether any of them has more weight than another or whether they "team up" and jointly exert influence on the structure of the lexicon. At any rate, when all factors are conflated into a single index of complexity, the correlation with analyzability in the lexicon that is observed is very strong. Thus, taken together, the identified factors together jointly account for the behavior of the sampled languages, and by means of them, it is

possible to extrapolate from the sample on the entire population of languages presently spoken and to make some predictions about their behavior. Intra-family comparison revealed that often the same dependencies that are observed in inter-language comparison hold, that is, genealogically related languages are subject to the same trend. As a candidate for a functional motivation for the correlations, homonymy avoidance was discussed, though there are difficulties in demonstrating how precisely this putative principle operates, and it may be that a less strong, but more reliable, case can be made for a weak functional principle that balances off between phonological and lexical complexity, which are poles of a continuum on which languages place themselves somewhere along the axes.

As for predilections for either metaphor- or contiguity-based conceptualization of the investigated meanings, the main relevant factor turned out to be differences in the favored word-formation device. Languages with many derived terms favor, by the nature of the process, contiguity-driven relations in analyzable terms, while languages with more analyzable terms of the lexical type tend to have more metaphor-based denominations. There are other structural factors for which a statistically significant influence can be observed, but the functional connection of them to this variable are unclear. In addition, several smaller sections and excursuses, some of which have to be seen as preliminary investigations of an at times speculative nature, were devoted to topics such as analyzability in reconstructed proto-languages (with particular reference to Indo-European), to differential degrees of borrowing in languages of the Americas, and to differences between languages in the lexicon in noun- as opposed to verb-based orientation.

Chapter 6

Results II:

Semantic Associations and Their

Cross-linguistic Distribution

6.1. INTRODUCTION

Chapter 5 was concerned mostly with quantitative, statistics-based analysis of the behavior of individual languages with regard to the structure of their lexicon. This chapter, in contrast, is concerned with the semantic side of things, and investigates the patterns of associations found in motivated lexical items. § 6.2. systematizes the results for a selection of particularly noteworthy semantic fields and the ties between the meanings within them. It shows, inspired by Hjelmslev's (1963) strucuralist analysis of the organization with respect to the meanings 'tree,' 'wood' and 'forest' (compare also Haspelmath 2003: 237), the differences across languages in how they "carve up" the relevant semantic space, as well as some common metaphorical extensions (from 'eye,' 'mouth,' and 'faeces') to other entities not contigously related to them. § 6.3. then asks whether there is non-random variation in lexico-semantic associations depending on the climatic and geographical environment languages are spoken in. The extensive discussion in § 6.4. is concerned with yet other possible sources of non-random variation, namely the possibility of the spread of particular patterns within languages families by genealogical inheritance, as well as spread due to language contact and resulting areality with regard to semantic associations. Another concern of this section lies in globally recurrent and frequent patterns of lexico-semantic associations. This is the closest the present work comes to the *locus classicus* of linguistic typology: the hunt for universals. Furthermore, the section asks what, if anything, we can learn from these about cognition.

6.2. SEMANTIC FIELDS AND THEIR LEXICAL ORGANIZATION CROSS-LINGUISTICALLY

6.2.1. INTRODUCTION

Figure 1 is an adjacency network of lexico-semantic associations based on the entire database. It is based on an adjacency matrix of the lexical associations in the sample data, and plotted by using a visualization technique kindly computed by Michael Cysouw. In the network, the closer the meanings are associated with each other, that is, the more frequent the respective association is found in the database, the closer the meanings are to each other, and the shorter the branches connecting them. To make sure that the network remains readable, only associations found in more than nine languages of the sample are displayed. This is a constraint imposed by problems with readability: if all associations are displayed simultaneously, the diagram becomes unreadable. There is no reason inherent in the data why not the full network should be plotted.

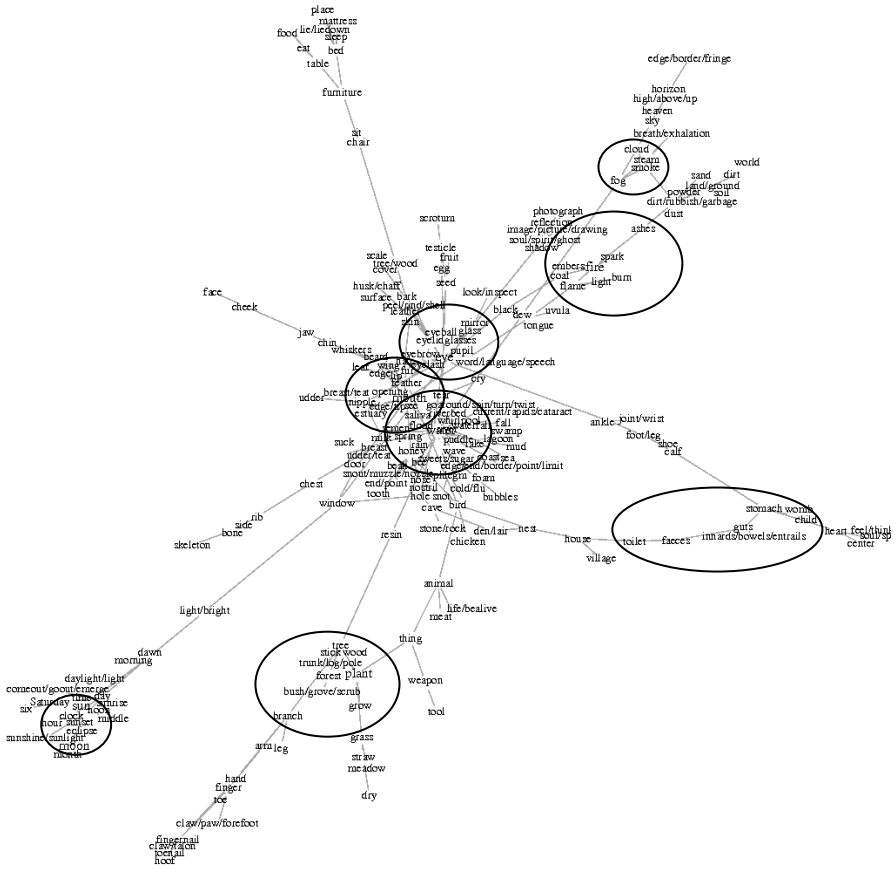


fig. 1.: an adjacency network of lexico-semantic associations

There are a number of meanings which occupy central positions in the network, while others are peripheral. This is not accidental, because the former are associated with a large number of concepts. Specifically, this is true of the heavenly bodies, water- and fire-related meanings, aerosols, and plants and their parts. The following paragraphs successively pick out one of these clusters and discuss in more detail how the meanings occupying central positions relate to the semantic fields surrounding them. At the same time, they point out differences in the lexical distinctions different languages make in a given semantic domain. In addition, further sections are devoted to internal organs of the body and body fluids; many of them do not show up in the diagram because the ties between them are relatively weak, but they showcase interesting interrelations.

Furthermore, there are four meanings that figure prominently in the network not because they are related to a large number of meanings contigously, but because they are frequent source concepts in metaphor-driven denominations for a large number of meanings from a wide variety of semantic domains. These are 'eye,' 'mouth,' 'faeces,' and kinship terms. Given that the network only shows associations occurring nine or more times in the languages of the sample, these associations are only hinted at there: note, for instance, the proximity of 'eye' to 'spring' and of 'mouth' to 'estuary.' More thorough discussion in the relevant paragraphs underscores that metaphorical transfer of the aforementioned meanings is also found to other referents.

Dixon (2010: 256) maintains that "[f]or a study to qualify as lexical typology it should involve comparison of a tightly knit set of terms, the meaning of each being with respect to the meanings of the other terms in the set (just as in a grammatical system)," and the discussion to follow goes in that direction.

6.2.2. SEMANTIC FIELDS AND THEIR ORGANIZATION ACROSS LANGUAGES

6.2.2.1. The heavenly bodies

Figure 3 provides a diagrammatic representation of lexical associations between words for the heavenly bodies, that is, the 'sun,' the 'moon,' and the 'stars.'

The policy for this and all diagrams to follow is, in order to keep the discussion manageable, that only meanings figuring on the list of meanings on the original list are displayed (of course it would be possible to also include other associated meanings within the respective semantic domains as described in the discussion in Appendix E.). The thickness of the lines represents the strength of the association in the languages of the sample (the thicker the line between the boxes, the more languages exhibit an association between the meanings within the boxes). In addition, arrows indicate the direction of the mapping as revealed by morphologically complex terms: if a line has only an arrow in one direction, as that between 'sun' and 'moon' in figure 2, it means that the meaning 'moon' may be expressed by morphologically complex terms with one constituent being 'sun,' but not the other way around. Size of the arrows gives a rough idea of the prevalence of the mapping directions. Thus, a large arrow on one side of the line and a smaller arrow on the other indicates that the mapping is in both directions, but more frequent in one than the other. A thick line with small arrows on one or either side would indicate that the association is mostly by colexification, with some cases of realization by morphological complex-

ity, a line with no arrows at all means that the association is exclusively by colexification, etc.

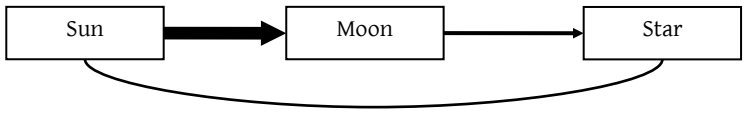


fig 2.: the heavenly bodies and lexical ties between their expressions

There is a common lexical relationship between 'sun' and 'moon' in languages of the Circum-Pacific language area as defined by Bickel and Nichols (in press), cf. (Urban 2009). Even in these languages, however, words for 'star' are virtually always distinct in the sense that they are either completely unrelated lexically and do not share morphological material, or that they are motivated by a complex expression based on the common word for both 'sun' and 'moon.' There is no language in the sample in which the former situation, colexification of all three referents, is found, and only one potential case of the latter situation: in Hupda, *wædhɔm'ǽh* 'star' might consist of *wædhɔ* 'sun, moon' and *mæh* 'small,' but this is considered unsure in the consulted source. Aside from languages of the Circum-Pacific area, it is the normal situation to have distinct words for 'sun' and 'moon,' and mostly also for 'star,' although in some cases the latter meaning may be expressed by complex terms deriving from either 'sun' or 'moon,' as is the case for instance in Guaraní and Wayampi. There is, however, one case of a language that expresses the meanings 'sun' and 'star' with the same word, namely the Australian language Burarra and also one case of a language with one term for 'star' and 'moon,' Abipón. Moreover, Bislama *sta* is glossed as "any heavenly body (e.g. moon, star, meteorite)." Table 1 provides examples of the lexical differentiation of the field in different languages (an asterisk after terms indicates that the language also features semantically more specific terms for one or more of the meanings colexified).

	Macaguán	Burarra	Abipón	Wayampi	Kosarek Yale
'moon'	*jomét, -omét**	*anjirderda, ran.gu*	*eergRaik*	*yaɨ*	*wal*
'sun'		*marrnga*	*mpaeRa, grahaolai*	*kwalaɨ*	*heng*
'star'	*jarwát*		*eergRaik*	*yaɨ-tata*	*douang, imbidea*

table 1: lexical differentiation for the heavenly bodies cross-linguistically

At any rate, the Burarra, Abipón, and Bislama cases appear to be extreme typological rarities judging from the evidence of the sample (it would be interesting to know if the situation that is encountered in Burarra has parallels in other Australian languages, although it does not appear to be too widespread).

6.2.2.2. Aerosols

Figure 3 represents cross-linguistic associations between terms for aerosols, that is, 'smoke,' 'steam,' 'cloud,' and 'fog' diagrammatically. As can be inferred from the figure, ties are quite strong and, with the exception of the pair 'smoke' – 'steam,' asymmetric when it comes to analyzable terms: complex terms for 'cloud' and 'fog' on the basis of 'smoke' are attested, as are complex terms for 'fog' based on 'cloud,' but not the other way around.

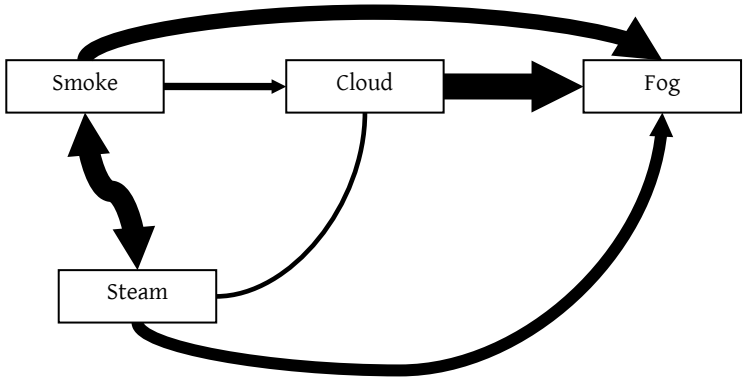


fig 3.: the semantic space of aerosols and lexical ties between its elements

There is just one candidate among the languages in the sample for employing a single term to cover the entire space of the semantic map of aerosols: the Zaparoan language Arabela, where *najaca* is used for all for meanings, though the complex expression *cohuaja najaca* (*cohuaja* means 'white') is (also) in use for 'cloud' and 'fog.' At the very least, it seems safe to say that all four referents contain the *najaca* element. Further, there are no morphologically unrelated synonyms listed for any of the four meanings in the source, which suggests that indeed *najaca* is the only conventional lexical expression associated with the meanings. Candidates for this type are also the Barbacoan language Cayapa and Tsafiki, where the lexemes *ñivijcha* and *poyó* respectively cover the meanings 'cloud,' 'smoke,' and 'steam.' They are candidates only because both sources do not cite the respective word for 'fog,' thus leaving open the possibility that this meaning is in fact expressed by a distinct lexical item. Tehuelche may be another case of a language with an at least likely diachronic relationship between terms for all four items in the semantic space: *p'aʔwn ~ p'awn ~ p'eʔwn ~ paʔwn* is synchronically 'cloud, fog,' while the phonologically very similar *p'aʔn ~ paːn* is used for 'smoke' and 'steam' (there is also the possibility that the *-w-* consonantism in the forms for 'cloud, fog' might be due to fossilized derivation by infixation diachronically).

Otherwise, in languages that cover three of the meanings with one single term, but employ a different one for the fourth, 'cloud' appears to be the one that is most commonly lexically distinguished. Examples are found in languages of Australia, more specifically Burarra and Yir Yoront. In Yir Yoront, *thorrqn* covers 'smoke,' 'steam,' and 'fog' (alongside 'haze' and 'spray from waves;' for 'smoke,' there is also a compound with *thum*

'fire'), while *yirrp* is used for 'cloud.' *Yirrp*, however, also means 'rain,' an instance of Australian "actual/potential-polysemy," and it is intriguing to speculate if this common Australian pattern contributed to the organization of this lexical field in Yir Yoront. Similarly, in Burarra, *jolnga* is used to refer to 'smoke,' 'haze,' 'vapor,' and 'fog,' while 'cloud' is *nguparr*, though note that there are competing unrelated synonyms or near-synonyms for some of the meanings. In contrast, in Anggor, the semantic range of *mburɨŋgai* is 'fog, mist, vapor, cloud,' while the lexically unrelated *hasahemɨ* is used to convey the meaning 'smoke.' There are no totally clear-cut instances of languages which treat 'smoke,' 'cloud,' and 'fog' lexically similarly, but 'steam' differently: in Maxakalí, *gõy* covers the three aforementioned meanings, but the source does not indicate how 'steam' is expressed (and in addition, there are compounds on the basis of *gõy* with *hãm*, reduced from *hahãm* 'land,' and *tex*, reduced from *tehex* 'rain,' for 'fog'). Furthermore, the Nez Perce lexical affix *ʔipé-* is glossed as 'pertaining to smoke, cloud, fog.' As the diagram in figure 4 also underscores, cross-linguistically, the ties between the meanings 'smoke' and 'steam' are more tightly knit than those with the other two meanings in the lexical field.

It is illuminating to move on to investigate whether there are languages which have two terms each of which cover two of the four meanings in the domain, because, surprisingly, such languages are quite rare. Next to the Tehuelche case already mentioned, Kwoma is an example of such a language; here, *hejagwayap* is used for 'cloud' and 'fog' and *hirika* for 'smoke' and 'steam.'

In contrast, languages which express two of the meanings in the semantic space by one term and the other two by unrelated terms are of course amply attested, though not in all possible configurations. In line with the stronger ties between the meanings 'smoke' and 'steam' on the one hand and 'cloud' and 'fog' on the other, one finds languages that use a single term for 'smoke' and 'steam' and unrelated ones for 'cloud' and 'fog' (Buli is an example) as well as the reverse situation expected from the general strength between the lexico-semantic connections, that is, languages that use the same word for 'cloud' and 'fog' but different ones for 'steam' and 'smoke' (Baruya is an example). There appears to be an ontology-based motivation for this situation: while both smoke and steam can be observed to emanate and rise up from elemental natural phenomena, fire and water respectively, neither fog nor clouds do, whilst there is an element of perceptual similarity to the latter meanings in that low clouds may appear similar to fog, and indeed the boundary, both meteorologically and perceptually, between the two is fluid to a certain degree. Examples of languages where the lexical organization of the domain in question cross-cuts this rather general division are less easy to find, but do exist: in Gurindji, for instance, *kaparru* means 'fog' and 'smoke-haze,' while *ngapurung* is used for 'steam' and 'fragrance from cooking' and *maarn* for 'cloud,' while in Kyaka, *popo* covers 'steam' and 'fog,' while there are unrelated items for 'smoke' and 'cloud,' and in Sedang, *kia hia* covers 'cloud,' 'smoke,' and 'air,' and there are different terms for 'steam' and 'fog.' Other combinations are not unambiguously attested. Table 2 summarizes the discussion and provides examples for each of the configurations mentioned.

	Ara-bela	Yir Yoront	Anggor	Buin	Kwoma	Buli	Baruya	Sedang	Kyaka	Gurindji
'smoke'	*(coh-uaja) najaca*	*thorrqn*	*hasa-hemɨ*	*iito*[1]	*hirika*	*nyuik*	*jɨta*	*kia hia**	*(isare) suk-wua*	*kapa-rru*
'steam'			*mbur-iŋgai**	*numa*			*mud-ɨnya*	*xoh, riôh*	*popo**	*nga-purung*
'fog'				*iito** *iito**	*hejag-wayap*	*koal-uk*	*yɨr-aaya*	*idrik, inoa*		*kappa-rru*
'cloud'		*yirrp*				*ching-mari*		*kia hia**	*kopa*	*maarn*

table 2: the semantic space of aerosols and different lexical configurations

Although the lexical associations recur on a global scale, there is nevertheless an areal hotspot in South America around the eastern slopes of the Andes.

6.2.2.3. Internal Organs of the Trunk

Lexico-semantic ties between the internal organs of the trunk, as visualized by the thin arrows between the boxes in figure 4, are on average relatively weak cross-linguistically.

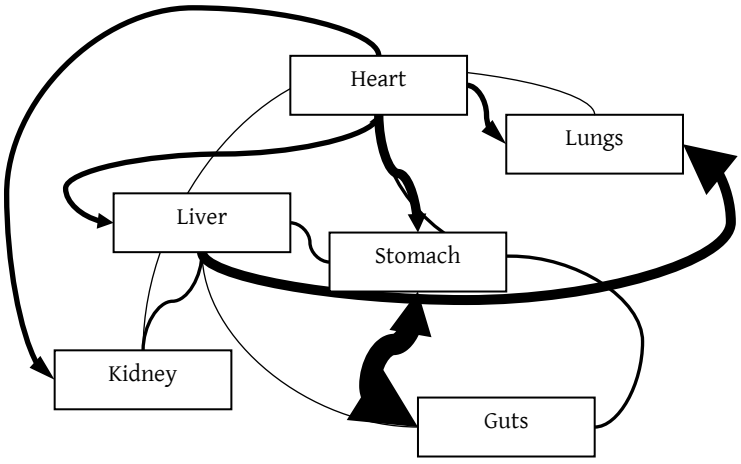


fig. 4.: internal organs of the body and ties between their lexical expressions

Perhaps the most noteworthy fact is that where such terms occur, they may connect a wide variety of the internal organs with each other (though not without restrictions, as will become clear later), and this fact is mirrored in diachrony in that terms for internal parts of the trunk are frequently subject to semantic change in which a term for one internal organ shifts to another. For instance, in Indo-Aryan, Vedic *vṛkká-* 'kidney' under-

[1] 'white cloud,' 'white smoke' more specifically and, according to the English-Buin finderlist, also 'fog.'

went semantic (and phonological) change to *bukkā-* 'heart' in a later stage of development. Still later, cognates came to mean 'belly' in Sinhalese, 'lungs' (among other meanings) in some Romani dialects, and came back full circle semantically to 'kidney' for instance in Tōrwāli (Turner 1966); see also Matisoff (1978) for some data from Tibeto-Burman.

The strongest connection is that between 'stomach' and 'guts,' which is unsurprising given the close spatial and functional proximity. Languages featuring a single term for 'stomach' and 'guts' are common, but complex terms also occur. Here, complex terms for 'guts' on the basis of 'stomach' (e.g. Toaripi *ére horou* 'belly rope') clearly outnumber complex terms for 'stomach' on the basis of 'guts' (e.g. Ngambay *kéy bò sìn* 'house big guts'). Relatively strong ties are also found between the meanings 'lungs' and 'liver' (compare Blust 2005 for Austronesian specifically). Colexification is attested for instance in Laz, but more often it is the case that the 'lungs' are expressed by a morphologically complex term based on the word for 'liver,' in which case the most frequent structures highlight the lesser weight of the lungs (e.g. Hawaiian *ake-māmā* 'liver-light') or their lighter color (e.g. Bislama *waet-leva* 'white/bright-liver'). While the 'liver' is thus often the source concept for the 'lungs,' so is the 'heart.' This is nicely illustrated by the terminological system for internal organs of the body found in the sampled varieties of Quechua, Ancash Quechua and Imbabura Quechua, shown in table 3.

	Ancash Quechua	Imbabura Quechua
'heart'	*shunqu ~ shonqu*	*shungu*
'liver'	*ñatin, yana ñatin* 'black liver'	*yana shungu* 'black heart'
'kidney'	*ruru-n* 'egg/testicle-3SG'	
'lungs'	*yuraq ñatin* 'white liver,' *yuraq shunqu* 'white heart'	*yurak shungu* 'white heart'

table 3.: terminology for internal organs of the body in varieties of Quechua

Both languages have an apparently cognate unanalyzable word for 'heart.' In Imbabura Quechua, this term serves to conceptualize other internal organs of the body: both 'liver' and 'kidney' are *yana shungu* 'black heart,' while the lungs are in contrast *yurak shungu* 'white heart.' The same structure for the latter meaning is found in Ancash Quechua, too. In fact, Imbabura Quechua is the language in the sample with the strongest lexical relationships between terms for internal organs of the body. In contrast, Ancash Quechua also uses the 'liver' as a conceptualization source: the lungs can also be called *yuraq ñatin* 'white liver,' and the meaning of the simplex *ñatin* can be reinforced and contrasted to the word for 'lungs' by *yana ñatin* 'black liver.'

Interestingly, the same situation -one meaning being the conceptualization target of both 'liver' and 'heart'- also pertains to the 'stomach,' although the ties are weaker in this case, and the data are not entirely straightforward. In Yanomámi, *amo* means 'piece of liver' (as well as 'central part of a plant' inter alia). The meaning 'liver' itself is rendered by *amoko* or *amokɨ* (*-ko* is a plural suffix and *kɨ* a quantal classifier, see § 4.4.1.1.), with *amoko* also being capable to refer to the 'stomach.' The situation with regard to the meanings 'heart' and 'stomach' is more straightforward: many languages colexify the meanings (for

instance Yuki, in which both meanings can be expressed by *ʈu·*), and in two languages, Kiowa and Malagasy, complex terms are found for 'stomach' (the Kiowa term, for instance, is *t'ę̨ɲ-t'o̜ų* 'heart-water;' when questioning consultants, the lexicographer received the reply that it is so called "because it is the place that the vomit comes from"). Note that both 'mouth' and 'stomach' are part of the digestive system and hence also contigously associated (compare the diachronic connection of Greek *stómachos* 'stomach' with *stóma* 'mouth'). [2] However, with regard to the source concepts 'heart' and 'liver,' there is no discernible directional pattern evidenced by analyzable terms. In Mbum, the 'heart' is *làù wârké* 'liver male,' while, as already seen, in Imbabura Quechua the 'liver' is *yana shungu* 'black heart.'

There is one internal organ of the body that stands out in that lexical ties with other internal organs are comparably weak cross-linguistically. These are the 'kidneys.' Colexification is found in three sampled languages with 'heart,' and also in three languages, complex terms for the 'kidneys' on the basis of 'heart' are found, one of them Imbabura Quechua. Colexification with 'liver' is found in Badaga (although the relevant term also conflates the meanings 'larynx' and 'lungs'), and there is one language, Kiowa, in which the same root *tə̀dl* is used for 'kidney' and 'liver' exclusively. However, a number of morphologically complex structures exist to disambiguate, among them *t̂adl-syᵸn* 'liver/kidney-small' = 'kidney' and *tə̀dl-eidl* 'liver/kidney-be.large' = 'liver.' The fact that cross-linguistically more complex structures exist for 'kidneys' than for 'liver' is suggestive that it is the latter meaning which is dominant and lexically more entrenched. Notably, there are no particular lexical ties of 'kidney,' 'lungs,' or 'stomach' in the sample, apart from cases in which one term has broad reference over a wide range of internal organs to be discussed now.

Sometimes, languages cover more than two points of the semantic space regarding the internal organs with one single term, and there appear to be few if any restrictions as to which organs can and cannot be so treated. For instance, in Ngambay, *wùr* may refer to 'liver,' 'belly' and 'heart' (and has a figurative meaning 'patience'). In Kwoma, the most salient meaning of *wopu* is 'liver,' judging from the microstructure of the dictionary entry, but a secondary reading is "vital organs generally (e.g. liver, heart, lungs)." An even more extreme case is presented by Khoekhoe *ǃnāb*, which means 'belly, stomach' in a narrow sense but also "innards, offal (i.e. lungs, heart, liver, kidneys)" in a more general sense. This is, next to the situation in Badaga already discussed above, the only case in which a term with such broad reference also includes the 'kidneys,' and it seems to be the only possible generalization that inclusion of this meaning in colexification is rare.

At any rate, it is conceivable that this case of synchronic colexification is a snapshot of an ongoing process of semantic generalization (as noted above, semantics of terms for internal organs of the body seem to be quite unstable). Interestingly, the apparent dominant reading is different in each case: 'liver' in Kwoma and 'belly/stomach' in Khoekhoe, which can be read as evidence that the starting point of the generalization is not

[2] In earlier language *stómachos* denoted 'throat,' 'gullet' and 'mouth (of the bladder, uterus)' and assumed the meaning '(upper orifice of the) stomach' later (Beekes 2010: 1408).

necessarily one particularly salient internal organ, but can be constituted by several different ones. Corroborating this, the synchronic cases of colexification involving a narrow and a more general reading have a diachronic correlate for instance in Indo-Aryan: Classical Sanskrit *phupphusa-* 'lungs' (that is, again a different internal organ) is continued in Sindhī as *papuvā* with the same meaning, but in the plural form *papu*, it refers to 'heart and liver and lungs, breast, bosom' (Turner 1966). Table 4 gives examples of cross-linguistic patterns of colexification in this semantic field (again, an asterisk indicates presence of more specific alternative terms), with Greek illustrating a language with maximal differentiation.

	Khoekhoe	Ngambay	Badaga	Laz	...	Greek
'heart'	*!nāb**	*wùr**	*karu**	*guri*		*kardiá*
'stomach'		*wùr**				*stómachi*
'lungs'		*pùpú*	*cuṭṭage ~ suṭṭage**	*cigeri*		*pnéumōn*
'liver'		*wùr**				*sykṓti, ἧpar*
'kidney'		*mùnjù*		*n/a*		*nefró, nefrós*

table 4: internal organs of the trunk and cross-linguistic patterns of colexification

An obvious question that arises is whether the patterns in the linguistic treatment of the internal organs of the body have any physiological grounding, that is, whether they can be explained by the perceptual properties of the organs. This is most clearly the case for the lexical connections between 'lungs' and 'liver': they are situated in close spatial proximity in the human body; they are both big organs, but differ in color (the liver is reddish brown while the lungs are pink) and in weight (the liver is the heaviest internal organ of humans, which nicely explains the conceptualizations mentioned above). In general, lexico-semantic ties are strongest for the four organs positioned roughly in the center of the trunk: the 'heart,' the 'liver,' the 'lungs,' and the 'stomach' (although the latter has for obvious reasons also pronounced connections with 'guts' cross-linguistically). Thus position within the trunk seems to be one explanatory dimension. Together with a second dimension, that of size, an even more complete picture emerges. Given that there is a hierarchy between the organs with respect to size (liver > lungs > heart > stomach > kidney), one can explain the strong ties between 'liver' and 'lungs' on the one hand, as well as the relatively strong ties between 'heart' and 'stomach,' in particular by colexification, on the other. Note that organs on the endpoints of the hierarchy tend to show few connections. This is true of 'liver' and 'stomach' (in spite of anatomical proximity), but it is particularly conspicuous with respect to the linguistic recognition of the 'kidneys': their peripheral position as well as their small size explain the paucity of lexico-semantic ties with other organs (in addition, they have a notably pronounced shape, and thus motivated terms in many languages make reference to that rather than to other internal organs of the body, see Appendix E, 129 for full discussion).

6.2.2.4. Body fluids

As can be seen from the diagram in figure 5, there are comparably weak lexical ties between the terms for body fluids cross-linguistically. However, these ties exist interestingly between the majority of the individual body fluids and are not, as one might suspect, restricted to just a few of them while others are completely unconnected in all sampled languages. These are diagnosed as being metaphorical in the present framework (compare Rice's to appear metaphor EFFLUVIA ARE OTHER EFFLUVIA to account for such semantic associations in Dene Sųłiné).

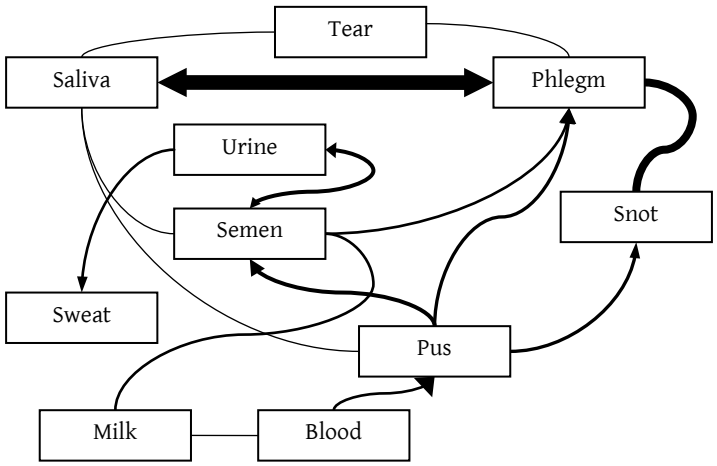


fig. 5.: body fluids and ties between their lexical expressions

Among the strongest ties in the above diagram are those connecting the meanings 'saliva,' 'phlegm,' and 'snot,' that is, those body fluids that have their origin in parts of the respiratory system. Perhaps the most surprising finding is the relatively central role that the concept 'pus' plays: complex terms for 'semen' and 'snot' are in fact found at times on the basis of 'pus,' for instance, Abzakh Adyghe has *pe-šən* 'nose-pus' for 'viscous snot' and Nez Perce *simqéheqs* /simqé-heqes/ 'penis-pus' for 'semen,' but the other logically possible direction is unattested in the sample. However, terms for 'pus' that are secondary to those for 'blood,' like Tetun *raan-kroek* 'blood-rotten' and *raan-mutin* 'blood-white' are attested.

Otherwise, the distribution of the associations allow for little systematization, which is not the least due to the fact that most of them are only attested in one language ('milk' – 'blood' in Kwoma, 'tears' – 'mucus' – 'spittle' in Khalkha, which also has distinct words for the two latter meanings, 'spit' – 'semen' in Rotokas, 'saliva, spittle' – 'pus' in Sedang, as well as perhaps the expression of the meaning 'sweat' on the basis of 'urine' in Guaraní). The association between 'milk' and 'semen' occurs in two languages, but here perceptual similarity in color is available as a hypothesis for the motivation of the association, and that with 'urine' occurs in two languages: by colexification in Tuscarora and by an archaic derived term for 'semen' from a verb meaning 'to urinate' in Khoekhoe.

The overall lesson to learn from the sample data is that lexical connections between body fluids are relatively rare (that is, they are referents most of the time expressed by unrelated lexical items), but they do occur, and there appears to be no general constraint as to what names for body fluids are particularly prone to be lexically associated. Such an apparently relatively unconstrained situation may be the outcome of taboos or, in a less strong form, by euphemistic meaning extension of terms for more "harmless" body fluids to more delicate ones. For diachrony, then, the upshot is that it is at least not impossible that terms for body fluids shift in meaning to other body fluids without any apparent semantically-based explanation for the shift other than that both referents belong to the same semantic domain.

6.2.2.5. Configurations of water

As is obvious from the diagrammatic representation in figure 6, 'water' plays a central role in this semantic field.

However, there are many more observations to be made about the organization of the field in individual languages. There are languages in the sample with no clear areal distribution in which 'water' and at least one major type of body of water, that is 'river' or 'lake,' and are not lexically distinguished at all (such a system is described in detail in Burenhult 2008b for Jahai and was alluded to earlier). The most extreme and unique case in the sample is Jarawara, which uses a single lexical item, *faha*, not only for 'water,' 'river,' and 'lake,' but also for 'rain' (there is also the word *isi/iso* for 'rain,' which also means 'leg,' 'handle' and 'stalk;' the only lexical alternative available for 'lake' is *rako*, which is a loanword from Portuguese).

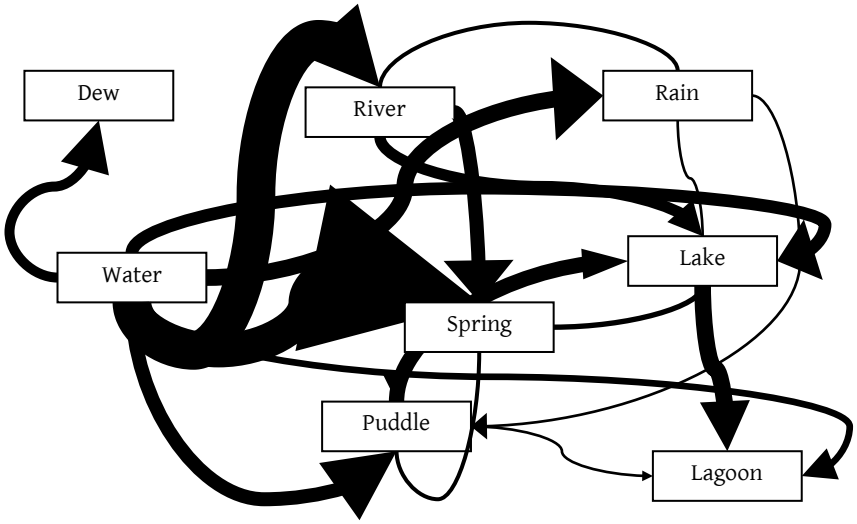


fig. 6.: terms for configurations of water and lexical ties between them

While typically languages with a system similar to Jahai employ different and unrelated terms for 'rain,' and, conversely, languages which use the same word for 'water' and 'rain' usually have unrelated or at least morphologically complex expressions for the different bodies of water, some languages with an overlap exist. The meaning of Waris *po*, for instance, ranges over the referents 'water,' 'river,' and 'rain,' while 'lake' is treated differently linguistically (the term might be a compound of *po* and the Waris word for 'son;' a similar system is found in Bakueri, where 'lake' is literally 'sea child'). The same situation is found in Kosarek Yale, also spoken in New Guinea. In contrast, in Berik, *fo* ranges semantically over 'water,' 'river,' and 'lake,' while 'rain' is expressed by the lexically unrelated *aro*, and in Itzaj *ja'* means 'water,' 'rain,' and 'lake,' but not 'river,' for which a number of other lexical labels exist (among them *ok ja'*, literally 'foot/leg water;' there are also formally redundant complex terms on the basis of *ja'* for 'lake'). Systems like these, in which two different configurations of water are denoted by a single lexical item which is also the designation for the substance 'water' are relatively rare cross-linguistically. The semantic field is somewhat more differentiated in Hup: here one encounters the same word denoting both 'water' and 'rain,' but different lexical expression for 'river' and 'lake' (one of the Hupda terms for 'river' is *deh-mí* 'water-waterway'). This system is fairly common cross-linguistically, but is particularly frequent in the Americas. The mirror-image of Hupda is Quileute, in which an even more common system is found. In languages of this type, the same monomorphemic lexical item is used to denote the substance 'water' and 'river,' while 'rain' and 'lake' are lexically differentiated. A conceivable situation is also one in which 'water' and 'lake' are colexified, but 'river' and 'rain' are treated differently linguistically. Comanche is the closest in the sample to that: *ɯmahpaa?* means 'rainwater' and 'pond, lake,' while the substance 'water' is *paa*, 'rain' is *ɯmapɨ̲* and '(small) river' is *okwèetɨ̲*. Not in all languages which lack lexical differentiation for different bodies of water is it necessarily the case that the term covering them is always at the same time expressing the meaning 'water.' For instance, Khalkha has a single term, *møren,* which may refer to both 'river' and 'lake,' and Bakueri has a single term, *mo̲rô̲*, for both 'river' and 'spring' while the substance 'water' is designated by the unrelated *málíwá*. The different systems are summarized in table 5, with Kildin Saami illustrating full lexical differentiation.

	Jarawara	Waris	Berik	Itzaj	Hupda	Quileute	Khalkha	Kildin Saami
'water'	*faha**	*po*	*fo*	*ja'*	*dĕh*	*kʷá·ya*	*usun*	*čāʒ'*
'rain'			*aro*	*ja'*		*łibó·kʷ̣*	*boruɣa(n), xura*	*ābb'r*
'river'			*fo*	*b'ekan, riiyoj, ok ja'*	*dehmí, má*	*kʷá·ya**	*møren**	*jōgk*
'lake'		*polomb*		*ja' (noj-ja')*	*móh*	*łtó·kʷ̣oł*		*jāvv'r*

table 5.: configurations of water and cross-linguistic patterns of colexification

The concept 'dew' is only connected to the semantic field by analyzable terms, and there is only one language in the sample, Nez Perce, which colexifies 'dew' and 'water' directly.

6.2.2.6. Fire and associated meanings

Similar to the semantic field of configurations of water discussed above, where terms for 'water' occupy a central position, the domain of fire-related concepts is obviously organized around 'fire,' as seen in figure 7.

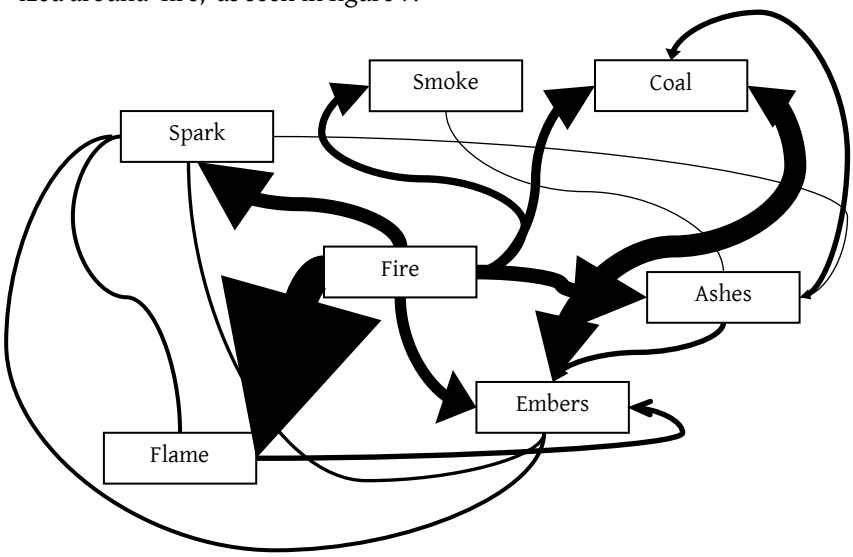


fig. 7.: terms for fire-related meanings and lexical ties between them

A language in which fire-related meanings are consistently contiguity-anchored is Toaripi. The majority of the relevant terms, as seen in table 6, however, are not fully analyzable on the basis of the consulted source.

Meaning	Toaripi equivalent
'fire'	*a*
'flame'	*a-uri* 'fire-tongue'
'spark'	*a-e* 'fire-faeces'
'ashes'	*a-futae* 'fire-??'
'embers'	*a-koela* 'fire-??'
'smoke'	*a-ikaera, a-ikoeila, a-ivuka, a ikohela* 'fire-??'
'coal'	*a-ro* 'fire-??'

table 6.: Toaripi terms for fire-related meanings.

Within this semantic field, however, colexification is relatively rare, which in all likelihood has something to do with the disparate perceptual properties of the referents. The most common pattern of colexification is that of 'flame' with 'fire' itself, followed by that with 'embers' and 'spark,' which are both relatively weak, however. Stronger associations

by colexification are found again between the meanings 'embers' and 'ashes,' as well as between 'embers' and 'coal' and 'ashes' and 'coal.' This is most likely due to the obvious fact that 'coal' and 'ashes' have a common semantic denominator: they are remnants of a burning fire. Rather than asking which meanings may be colexified, perhaps a more interesting question in this particular case is to ask which patterns of colexification are actually not attested. As already implied in the above discussion, the meanings in the above diagram essentially form two clusters for which there is an ontological basis: colexification is attested for meanings having to do with an actually burning fire ('flame,' 'spark') and those that are remants of a once burning fire ('ashes,' 'coal'), with 'embers' occupying an intermediate position and forming a link between the two clusters that mirrors its intermediate position in the process of a fire burning down.

6.2.2.7. Plants and their parts

Strongest ties in the semantic field of plants and their parts, as can be inferred from figure 8, are found for meronyms of 'tree,' and this is true for both colexification and morphologically complex terms.

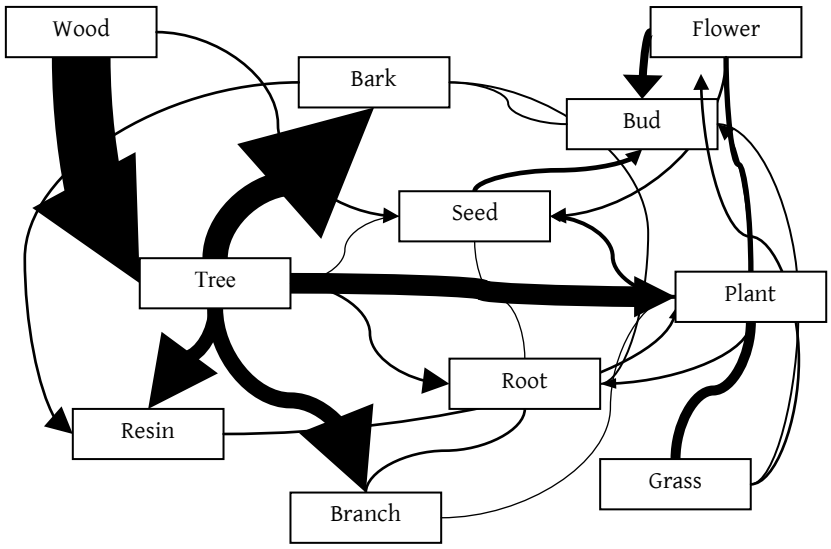


fig. 8.: terms for plants and their parts and lexical ties between them

In particular, colexification of 'tree' and 'wood' is frequent. In Witkowski et al.'s (1981) sample, colexification occurs in two thirds of sampled languages, and the figure obtained in the present study draws close to this value. As also noted by Witkowski et al. (1981: 5), there is a tendency for morphologically complex terms for 'tree' to be based on 'wood,' although in the data underlying this study, this pattern is not as strong as discussion in Witkowski et al. (1981) suggests. In fact, this situation is only found unambigously in one sampled language, Manange (which has *2ʃiŋ-3tuŋ* 'wood-copse/trunk), though interest-

ingly Lesser Antillean Creole French has a similar pattern (*pié-bwa* 'wood-stem'). In addition, the Cubeo term for 'tree' is *jocʉ-cʉ*, which is the term for 'wood,' *jocʉ*, suffixed with the classifier *-cʉ* for tree-like objects and furthermore, Upper Chehalis *ƛíš-aƛ'š* 'a clump of trees' is a reduplication of the term for 'wood.' Whether the reverse relationship also occurs and if so, how frequent it is, cannot presently be assessed since 'wood' itself is not among the items under investigation. Colexification of 'tree,' with 'trunk,' 'pole,' or 'log' is found in eighteen sampled languages.

As a comparison with those languages in which the colexification is with 'tree' makes clear, there is some overlap between the groups. Ngambay and Khalkha have single terms for both 'tree,' 'wood,' and 'forest' (and in the case of Ngambay, even 'branch') due to the extremely frequent colexification of 'tree' and 'wood' themselves (see Appendix E, 65), but there is also a relatively large number of languages which do not have this pattern of colexification, instead either having distinct terms for 'wood' and 'tree' and colexifying 'forest' with one of them. Table 7 provides examples of different patterns.

	Ngambay	Khalkha		Waris	Yaqui	Imbabura Quechua
'wood'	*kake**	*modu(n)**		*ti*	*kuta*	*kaspi*
'tree'					*juya*	*yura*
'forest'				*sungeit, ekla*		*sacha*
'branch'		*gesigyn*	~	*klal, tikla*	*buja*	*malki*
		*gesigyy**				

table 7. terms for parts of plants and cross-linguistic patterns of colexification

In some sampled languages, 'forest' is expressed by morphologically complex terms on the basis of 'tree,' as in Baruya *yɨ'darya*, literally 'tree area' and Ancash Quechua *sacha marka* 'tree/plant area' (see Appendix E, 26 for more details). An interesting variation of complex terms of the Baruya and Ancash Quechua type is found in three languages of South America, Jarawara, Lengua, and Yanomámi. Here, the general meaning 'place' is colexified with 'forest' (on semantically general terms of this kind in South America and Jarawara specifically, see § 6.4.3.15.). This is also interesting in the light of the impact of environmental factors on the colexification of particular meanings, since at least Jarawara and Yanomámi are spoken in the tropical rainforest of the Amazon basin.

Moreover, meronyms of 'tree' are commonly expressed by morphologically complex terms on the basis of a term with just that meaning. A language in which 'bark,' 'resin,' and 'branch' are all expressed by morphologically complex terms is San Mateo del Mar Huave; in fact it is the only sampled language in which all three meanings are expressed by analyzable terms of which 'tree' is one constituent (see table 8), but the semantic relations between these are not very frequent cross-linguistically (see discussion of individual concepts in the relevant sections Appendix E).

Meaning	Sam Mateo del Mar Huave Equivalent
'bark'	*mipang xiül* 'shell tree'
'resin'	*aonts xiül* 'excrete tree'
'branch'	*omal xiül* 'point tree'

table 8: San Mateo del Mar Huave terms for meronyms of 'tree'

Of these, 'resin' may also be based on 'bark,' as in Piro *mta-ha* 'bark-water.'

Apart from meronyms of 'tree,' the lexico-semantic ties are cross-linguistically relatively weak, but one less tight cluster is discernible which consists of meanings having to do with the reproductive system of plants: 'bud,' 'flower,' 'seed' (and 'fruit'). Obviously, the 'flower,' 'bud,' and 'fruit' stand in a relationship of temporal contiguity with each other, and, in addition, 'seed' stands in a meronymic relationship with 'fruit.' Within this field, ties between 'bud' and 'flower' and 'seed' and 'fruit' respectively are particularly strong (relatively speaking). For instance, in Wayampi, 'bud' is *pɔti-yaʔɨ* 'flower-child' (which is also an interesting denomination because of the metaphorical transfer of 'child,' having to do with reproduction in humans or animates more generally, to the fauna). Very strong are the ties (found in seventeen languages) between the meanings 'seed' and 'fruit,'[3] but in spite of the obvious contiguous relationship between 'flower,' 'bud' and 'fruit,' colexification of 'bud' with 'fruit' or 'flower' is comparatively rare. In Kaluli, the meaning colexified with 'flower' is more precisely 'inedible tree fruit,' and it is this fact which points to a possible explanation of the observed frequencies. 'Flowers' and 'buds' are of no or quite limited use for humans, while 'fruits' are in that some of them are edible, and thus their quality to potentially serve as foodstuff may be an important component of the lexical semantics of terms for 'fruit' cross-linguistically. Thus, it is not surprising that colexification of the kind mentioned above is relatively rare, in that 'fruits,' in terms of Gibson (1979), are likely to be conceptualized under the perspective of human affordance, whereas 'flower' and 'bud' are not. Table 9 provides an overview over the elaborateness of lexical differentiation for the meanings just discussed.

	Efik	Sahu	Lesser Antillean Creole French	Baruya	Kiliwa
'fruit'	*m'fri* ~ *mfuri**	*palingasa**	*n/a*	*n/a*	*tkwma?, -pay*
'flower'			*flé*	*purɨrya*	*tpyawp*
'seed'		*moi'i*	*jem**	*wia*	*tyit*
'bud'		*boro*		*purɨrya*	*chiilp*

table 9.: terms for some parts of plants and cross-linguistic patterns of colexification

[3] The situation in Toaripi is also discussed in Brown's (1972: 171) comparative semantic analysis of Toaripi and the related Orokolo. He states that "while the term *fare/hae* covers both the meanings 'seed' and 'fruit', with fleshy types of fruit it has reference rather to the seed or nut, and not to the fruit as a whole."

Among the other minor patterns not made explicit in figure 9 since not all of the respective meanings figure on the original meaning list are colexification of 'leaf' and 'branch' in Gurindji and Nuuchahnulth, which can be explained by the spatial contiguity between the two meanings, colexification of 'flower' and 'pod' in Rao, of 'leaf' and 'flower' in Cheyenne, and that of 'bud' and 'young leaf' in Efik, Sko, Jarawara and Lesser Antillean Freole French.

6.2.3. BODY-PART METAPHORS

6.2.3.1. The Eye

The 'eye,' as the most salient feature of the human face (Shepherd et al. 1981), is an extremely common conceptualization source for a wide range of meanings, including both other, presumably less salient body parts, and many meanings in other semantic domains. There is also literature on this for Austronesian languages specifically (Barnes 1977, Chowning 1996), but this type of transfer is common across the globe. The diagram in figure 9 provides an overview of lexico-semantic associations with 'eye' (dashed lines indicate that the association is only present by semianalyzable terms in this and further diagrams in the following two sections).

In the upper left corner, there are contiguity-based conceptualizations on the basis of 'eye' for body-parts that are immediately adjacent to the eye, or are more properly put parts of it, as well as for 'tear,' a body-fluid that is in contiguity with the 'eye' since this is where it originates. From a conceptual point of view, these are fairly uninteresting. What is noteworthy, though, is the large number of languages, as indicated by the thick black arrows, in which 'eyebrow,' 'eyelash,' 'eyelid,' 'pupil,' and 'tear' are transparent complex expressions based on the respective terms for 'eye.'

A more interesting question, however, is whether there is an all-or nothing situation, that is, whether languages either favor having complex lexemes for all of the body-parts in contiguous association with the eye or to have unanalyzable terms for the entire set of meanings. The answer is that there is little evidence for such a principled linguistic treatment. There are languages in which terms for the entire set of contiguously related meanings are analyzable (as summarized in table 10 for Kashaya, which is such a language), but a more frequent situation is that languages fall somewhere in between, with some analyzable terms and some unanalyzable ones.

Lexical Item	Underlying Representation and Gloss	Meaning
huʔu· sime	/huʔuy sime/ 'eye fur'	'eyebrow'
huʔu· pitemʔ[4]	/huʔuy pitemʔ/ 'eye droop.of.eyes'	'eyelash'
huʔuy šiʔda	/huʔuy šiʔda/ 'eye skin'	'eyelid'
huʔu· qʰaʔbe	/huʔuy qʰaʔbe/ 'eye rock'	'eyeball'
ʔu·qʰa[5]	/huʔuy ahqʰa/ 'eye water'	'tear'
huʔu· šihta	/huʔuy šihta/ 'eye bird'	'pupil'

table 10.: analyzable items for meanings related to 'eye' in Kashaya

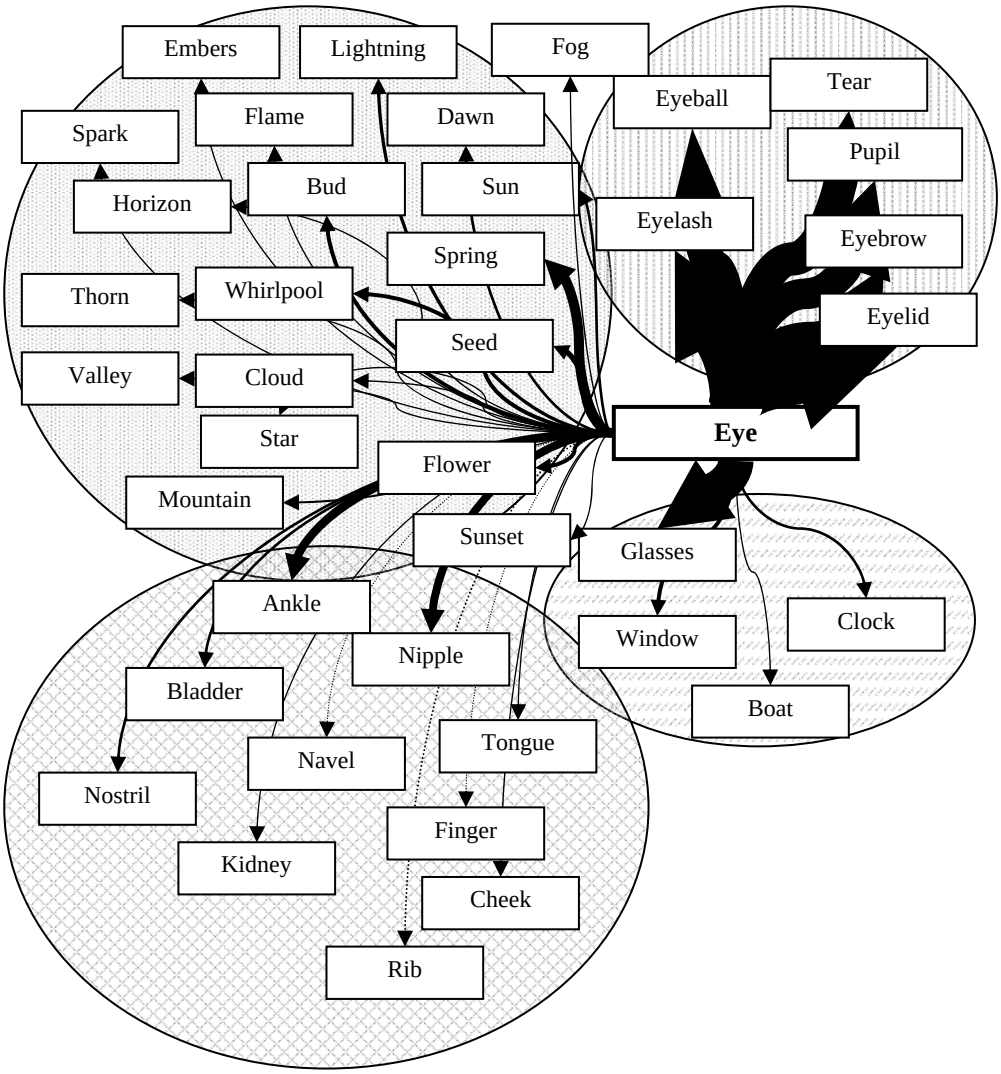


fig. 9.: lexico-semantic associations for the 'eye'

[4] *huʔu· sime* may be used as well.

[5] Glossed as 'eye water' in the consulted source, but possibly lexicalized.

In the lower right corner of the diagram in figure 9, there are a variety of artifact terms which cross-linguistically are sometimes conceptualized via 'eye.' In the case of 'eyeglasses,' this is obviously due to contiguity, while in the case of 'clock,' it is due to the similarity in shape between the two referents, perhaps aided by the additional similarity in shape with the 'sun.' While the presence of the word for 'eye' in the Yoruba word for 'boat,' *ọkọ̀-oju-emi* 'vehicle-eye-water' remains unclear, similarity in shape is also available as an explanation for the presence of the words for 'eye' in terms for 'needle' and 'window.' Similarity in roundish shape is also a likely factor explaining the conceptualization of 'windows' via 'eye' in many languages (and note the etymology of English *window*, which, according to the Oxford English Dictionary, goes back to Old Norse *vindauga*, a compound of *vindr* 'wind' and *auga* 'eye,' as well as the precise parallel noted in Yoruba in table 11).

In fact, discussion of the case of 'window' provides a good transition to the many meanings in the domains of nature-related and body-part terms that may be expressed cross-linguistically by analyzable terms on the basis of 'eye.' The similarity in roundish shape explains most of the associations that occur with body-part terms. The 'ankle,' the 'nipples,' the 'navel,' the 'bladder,' the 'kidney,' and 'the nostrils' quite obviously are round in shape, and many of them are also roughly comparable in size to the 'eye.' More remotely roundish are the 'tongue,' the 'cheeks,' the 'ribs,' and the 'finger.' In these cases, it can be conjectured that there is another semantic feature in which these entities may be perceived as being similar to the 'eye,' namely three-dimensionality.

Similarity in round shape explains also the most frequent metaphor-driven transfers to terms for topological features, most prominently 'spring' and 'whirlpool' as well as those to 'seed,' 'bud,' and, from there on, 'flower.' (the associations with 'fog' in Kiliwa *yuw=hi?* 'eye-cover' and with 'horizon' in Khoekhoe *mūs lkhâu-s* 'eye radiate-3SG.FEM' appear to be contiguity-based). For two of the associations with meanings related to 'fire,' namely 'embers' end 'spark,' roundish shape may also be adduced as underlying the metaphorical transfer.[6] However, an additional component of brightness may well play a role, which would then also account for the associations with 'lightning' and 'dawn' and, most importantly, 'sun,' an association which is particularly common in languages of Southeast Asia and Oceania (Urban 2010). The association with 'dawn,' however, may also be explained with reference to shape-based similarity alone, given that the very first light of the day at dawn emerging from the horizon in fact is remotely roundish in shape.

The obviousness with which shape-based similarity is detectable is gradient. Thus, the associations with nature-related features such as 'mountain,' 'valley,' 'thorn,' and 'cloud' intuitively appears to be conceptually more remote than that with, say, 'seed.' However, all may be conceived on some level of abstraction as being roundish entities. Two additional remarks are in order: first, as mentioned above, the referents in question are not only round, but also three-dimensional objects, and this appears to be in some cases a secondary motivating factor. Second, the bolder conceptual transfers cluster in a

[6] The association with 'flame,' occurring in Kyaka, is a little less clear. However, *lenge*, which is the relevant term, has many meanings in Kyaka, among them 'node or knuckle,' 'stratum or narrow (vertical) panel in man's skirt-net,' 'woven body of a bag,' 'eye of boil or carbuncle' and last but not least 'eye of the head.'

relatively small number of languages, and it may well be the case that the presence of a number of the more obvious metaphorical patterns is a prerequisite for the development of conceptually less nearby transfers, with concomitant semantic bleaching and generalization of the term for 'eye.' Among the languages which have a variety of the latter is Yoruba, as seen in table 11.

Lexical Item	Gloss	Meaning
ojú ọ̀run	'eye heaven'	'cloud'
ojú-sanmà	'eye-sky'	'cloud'
ọkọ̀-oju-omi	'vehicle-eye-water'	'boat'
ojúaféfé	'eye-wind'	'window'

table 11: metaphor-driven complex lexical items in Yoruba involving *oju* 'eye'

Another language in which a variety of terms for natural kinds and artifacts are expressed using 'eye' as a source concept is Rama (Chibchan):

Lexical Item	Gloss	Meaning
king-úp, kung-úp[7]	'head/top-eye'	'mountain'
kat up	'tree eye'	'fruit, peanut'
kú up	'bird's wing eye'	'thorn, prickle'
ngústi úp	'pissing eye'	'bladder'
isúl-uk up	'??-skin eye'	'finger'
'píns-up ~ pínsh-uk	'?? eye'	'navel'

table 12: metaphor-driven complex lexical items in Rama involving *up* 'eye'

Valuably, Rigby and Schneider (1989) have at times included comments by their Rama consultant for lexical items. For *'píns-up ~ pínsh-uk* 'navel,' the consultant provided the literal meaning 'belly eye,' while in the entry for 'finger,' a comment on *up* as occurring in this term is 'something round.' This may have something to do with the term for 'finger' being only semianalyzable, but it is interesting for a general discussion of the semantic extension of 'eye' to round-shaped objects that here the semantic content 'eye' seems to be bleached to 'something round' when occurring in some complex constructions. This is in line both with the suggestion that *up* acts as a device for nominal classification (see § 4.4.1. for discussion) as well as with the observations to be made in § 6.2.3.3. in discussing extensions of 'faeces' or 'excretion': in some languages, there appear to be semantic templates on the basis of certain meanings to conceptualize a wide variety of referents on the basis of a bodily notion (see Levinson 1994 for discussion of body-part extensions in Tzeltal). These observations have obviously parallels to both grammaticalization and lexicalization: they are similar to grammaticalization in that a certain amount of semantic bleaching and generalization is involved, and to lexicalization in that the complex structures of this type may well not be transparent anymore to language users precisely be-

[7] There are more semianalyzable terms for 'mountain' where one of the constituents is *up* 'eye.'

cause of the semantic bleaching that seems to be a concomitant effect of an increasing use of the same source concept to name a large number of entities.

Summarizing, one can note three subtypes of metaphorical transfer of terms for 'eye' to other concepts, which, at least in the domain of body-parts, may receive additional support if the target concept is similar to the 'eye' in size.

(i) Shape-based similarity alone: roundness
(ii) Shape-based similarity with the putative additional motivating component of brightness
(iii) Shape based similarity with the putative additional component of three-dimensionality.

6.2.3.2. The Mouth

Figure 10 represents semantic associations for 'mouth' diagrammatically.

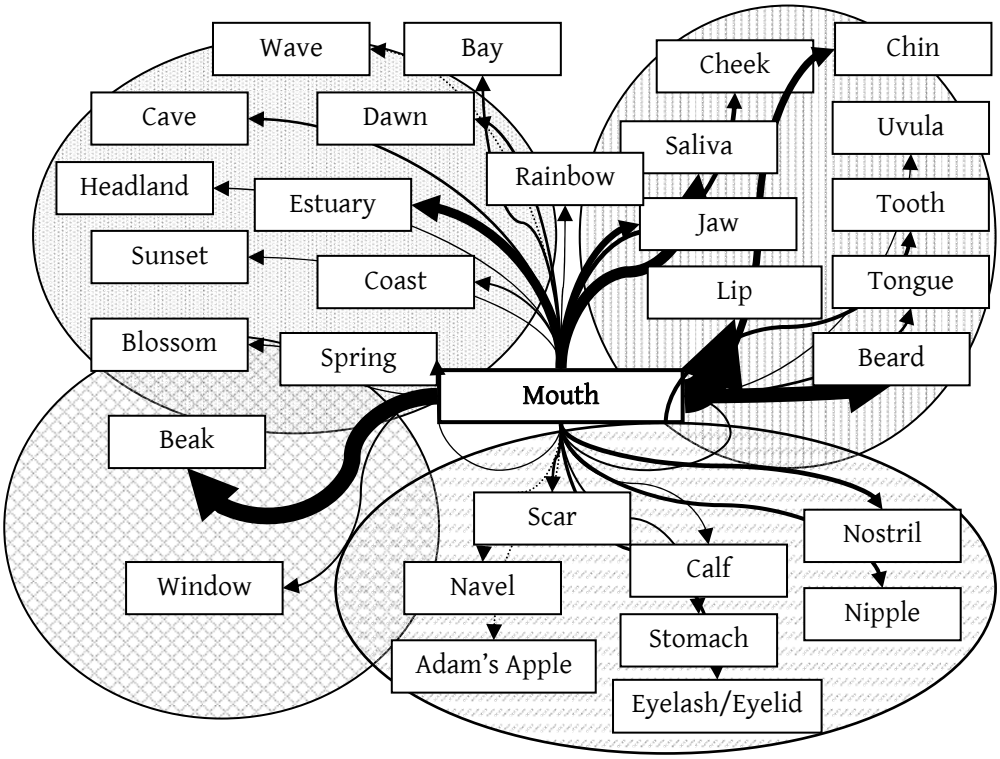


fig. 10.: semantic associations for 'mouth'

As in the above diagrams, concepts – facial features in this case – which are contiguously related to the 'mouth' are found in the top right corner. The situation here is comparable with that for 'eye', namely that the lexico-semantic ties in this area are strongest: terms for 'lips' frequently consist of the respective terms for 'mouth' and 'skin'

crosslinguistically, as in Cayapa *fi'pa'-quica* 'mouth-skin.' Terms for 'beard' which may be "literally" translated as 'mouth-hair', such as Wintu *qol-ċekey*, are common as well. 'Saliva' is frequently expressed by morphologically complex terms consisting of terms for 'mouth' and 'water' or 'liquid' more generally, such as Wappo *na-méy* (Appendix E, 133), and 'jaw'-terms are sometimes analyzable as complex structures involving 'mouth' and 'bone' (Appendix E, 118), as in Baruya *maanagɨnya*, which contains *maanga* 'mouth' and *yagɨnya* 'bone' and also means 'chin.' In one language, Aymara, 'tooth' is *laka ch'akha* 'mouth bone' (see Appendix E, 144 for more thorough discussion).

Sometimes the associations with topological features may exist because the word for 'mouth' also has a semantically bleached more general reading in the languages mentioned explicitly in the consulted sources. Often this is 'opening.' This is for instance the case in Bororo, one of the languages in which the respective term is present in the word for 'cave' (*ia-ri*, *ri* 'stone'), in Toaripi, where it is present in the word for 'nostril' (*ever-ape*, *ever* 'nose'). In Kiliwa, *ha?* in *miy=ha?* 'calf' also means 'face' alongside 'mouth,' and in Welsh *pen* in *pen-rhyn* 'headland' also means 'end, head, top.' Thus, one can basically observe the same pattern as with 'eye,' namely a certain amount of semantic generality and bleaching of the respective terms, in particular when occurring in morphologically complex expressions.

In spite of this, there is relatively little evidence for clustering of either contiguity-based analyzable terms for facial features or similarity-based extension to topological features of the environment in particular languages. A language where terms for facial features including the word for 'mouth' are frequent is Abzakh Adyghe, but here terms are often only semianalyzable (table 13):

Lexical Item	Gloss	Meaning
žač^{ʔ}e	/že-č^{ʔ}e/ 'mouth-end'	'beard'
že-pq^{ʔ}	/že-pq^{ʔ}(ə)/ 'mouth-skeleton'	'lower jaw, chin'
ʔ^{ə}ə-pś^{ʔ}e	'oral.cavity/opening -??'	'lip'
ʔ^{ə}ə-ps	/ʔ^{ə}ə-psə/ 'oral.cavity/opening-water'	'saliva'

table 13: contiguity-driven terms for facial features in Abzakh Adyghe

Two metaphor-based conceptualizations for the facial features 'beard' and 'chin, jaw,' with the term for 'mouth' acting as a contiguity anchor as defined in chapter 3, are found in Takia, as well as one for 'nostril' in which the 'mouth' serves as the source concepts' (table 14):

Lexical Item	Gloss	Meaning
awa-n dabi-n	'mouth-3SG root-3SG'	'beard'
awa-n to-n	'mouth-3SG arm-3SG'	'chin, jaw'
ŋdu-n awa-n	'nose-3SG mouth-3SG'	'nostril'

table 14: metaphor-driven terms for facial features in Takia

In contrast, pervasive conceptualization of topological features or other body-parts do not notably cluster in a particular language (at least not for the concepts presently looked at). Extensions of 'mouth' to nature-related and topological concepts are, comparable to the situation for 'eye,' often found when the target concept has roundish shape. This is true of virtually all of the meanings in this domain for which an association with 'mouth' is found, although, again, for the temporal concepts 'dawn' and 'sunset' the level of abstraction is somewhat higher than with the topological concepts. Unlike the patterns observed for 'eye,' however, objects which either have an opening, such as a 'cave,' or which involve the end or starting point of the passage of a substance or object (a figure) along some trajectory (here it seems useful to adopt terminology borrowed from Gestalt psychology into Cognitive Linguistics) appear to be a particularly amenable to being conceptualized via 'mouth.' This is most obviously the case for the associations between 'estuary' and 'mouth': just like the mouth is located at one end of the esophagus and respiratory tract, so the estuary constitutes one of the end points of a river.[8] But the patterning is also noticeable when it comes to the meanings 'sunset' and 'dawn' which relate to the beginning or endpoint of the movement of the sun in the sky. The case of the association with 'beak' is due to functional similarity, since the beak is the corresponding body part of birds to the mouth in that its functions include ingestion.

In the domain of human body-parts it is likewise roundness that appears to be the most prominent feature that triggers the fact that 'mouth' is used as a source concept, again obviously aided in the cases of the 'nostrils' and the 'stomach' by the fact that these body-parts have an opening (although the association with 'stomach' may be additionally supported by the contiguity between the two concepts as they both participate in the process of ingestion).

There are a number of concepts which can, judging from the evidence of the sample, cross-linguistically both be named with reference to 'eye' as well as to 'mouth': these include the 'nostril,' the 'nipple,' the 'window,' and the 'dawn' (and there are semianalyzable terms for 'navel' on the basis of 'eye' and 'mouth' in one language each). But notably, languages in which the word for 'nipple' is based on 'eye' are more frequent than those in which it is based on 'mouth.' Conversely, 'mouth' is a more frequent source concept for 'nostrils' than 'eye' is, and the same is true of 'dawn,' where associations with 'mouth' occur in two languages each but only in one with 'eye.' Thus, while their round shape makes them in principle amenable to being conceptualized both via 'eye' and 'mouth,' there appear to be some cross-linguistic preferences that have to do with whether the referents in question have openings or may be viewed as constituting parts of the trajectory of some objects. Size of the respective body-part term might play a role here, too: a 'nipple' is relatively small, comparable in size better to the 'eye' than to the 'mouth;' also note that 'eye' as a source concept for the slightly bigger 'calf of the leg' is not attested. For the case of the 'nostrils,' which are also comparatively small, the salience of their function as a trajectory (note also terms such as Baruya *si̶duta* /si̶nna-tuta/ 'path-

[8] The reverse naming pattern for 'estuary' is found in Jahai where *kit tɔm* 'river-mouth' is literally 'water-bum' (Burenhult 2008b: 186). Body-part metaphors for hydrological features are pervasive in this language.

nose/nostril') may override the factor size. Thus, tentatively: THE 'EYE' WILL BE A MORE FREQUENT SOURCE CONCEPTS FOR ROUND (THREE-DIMENSIONAL) OBJECTS, WHEREAS 'MOUTH' WILL BE MORE OFTEN UTILIZED AS A SOURCE CONCEPT WHEN THE TARGET CONCEPT IS ROUND, BUT WITH THE ADDITIONAL COMPONENT OF IT HAVING EITHER AN OPENING OR FORMING THE BEGINNING OR FINAL POINT OF THE TRAJECTORY OF SOME ENTITY.

There is one exception to the above generalization, namely 'window,' which is more frequently conceptualized via 'eye' than via 'mouth,' although a salient feature of windows is obviously that they are openings in the walls of houses. This is against what one would expect under the tentative generalization just made. However, there are other important factors that come into play here, namely that windows are functionally associated with seeing, as is the 'eye,' and that terms for 'window' may be embedded into a broader conceptual transfer pattern that likens human faces to houses, as in the examples in (1.) from Ma'di (Central Sudanic):

(1.) a. /dʒɔ́ tī/ 'house mouth' = 'door'
b./dʒɔ́ mī/ 'house eye' = 'window'
c. /dʒɔ́ drì/ 'house head' = 'roof'
(Blackings and Fabb 2003: 51, surface forms omitted)

Metaphorical denominations for 'door,' which are based on 'mouth' seem to be particularly frequent (see Zamponi 2009: 539 on Arawakan languages, Monod Bequelin 2006: 220 on Tzeltal), although the precise extent of this phenomenon cannot be assessed presently, since 'door' is not among the concepts on the wordlist under investigation here. And the passing-through function of doors is much more salient than that of windows (in addition to the fact that the latter have something to do with sight, which is another reason why 'eye' might be preferred as a source concept for 'window' cross-linguistically). This is probably why among the meronyms of 'house,' the 'door' rather than the 'window' are named using 'mouth' as the source concept, and the observations made with respect to 'door' and 'window' would then be readily accountable under the assumption of metaphors of 'mouth' for entities that are both roundish in shape and where some sort of spatial transition takes place. Summarizing the putative metaphoric transfer patterns, there are:

(i) Shape-based similarity alone: roundness
(ii) Shape-based similarity with the putative additional motivating component of target sources having an opening
(iii) Shape based similarity with the putative additional component of the target concepts forming the starting or ending point of a trajectory.

6.2.3.3. Faeces/Excretion

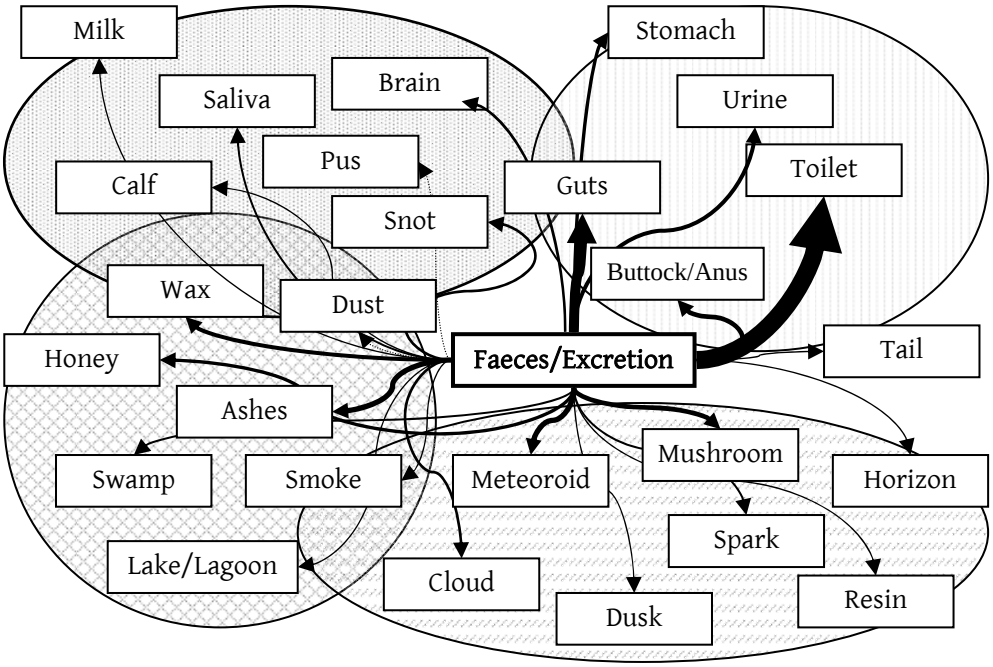


fig. 11: concepts with lexico-semantic associations to 'faeces' or 'excretion'

As represented diagrammatically in figure 11, lexico-semantic associations with 'faeces' may be grouped into four at times overlapping categories (not shown in the diagram is Bororo *baigabe* 'lightning,' which may be spurious: possible constituents are *baiga* 'type of Bororo bow' and *be* 'faeces'): on the one hand, in the top right corner are relatively obvious contiguity-based conceptualizations. These are noteworthy, but do not require extensive further discussion. In the top left corner other body-parts and body secretions are found, which are in some languages based on a semantic extension of 'faeces.' While there is a relatively obvious analogy between 'snot' and 'faeces' (this pattern is for instance found in Kashaya, where *ʔilahpʰa* 'dry snot' consists of *ʔila* 'nose' and *ahpʰa* 'excrement'), that between 'faeces' and 'brain' is somewhat less clear. Interestingly, this association occurs in the sample only in the Barbacoan languages Cayapa (*mishpe* /mishu-pe/ 'head-excrement') and Tsafiki (*fu-pe* 'hair-excrement). The Tsafiki word for 'wax' is also based on 'faeces.' What some of the body fluids in this group have in common is their undesirability.

The circle in the lower left corner groups together a number of referents where the association with 'faeces' is still fairly well motivatable in most cases, and which are clearly metaphorical in nature. 'Honey' and 'wax' are in a sense the most clearly perceivable correlates to excretions when it comes to bees, while 'ashes' are the (probably mostly useless, which can be construed as the tertium comparationis with 'faeces') remnants of a fire, and similarly, 'smoke' is emitted by a burning fire but is not readily utilizable in terms

of human affordance. The case of the association with 'lake' is fairly unclear. This comes from Cubeo, where *macajitabʉ* appears to consist of *maca* 'faeces' and *jitabʉ* 'puddle.'

The semantic associations in the lower right corner are the most interesting ones, because they tend to be highly abstract metaphors mostly. Table 15 lists languages with complex terms for 'mushroom' on the basis of 'faeces' (furthermore, Cashinahua colexifies these meanings directly).

Language	Term	Gloss
Rendille	*u'dú-yeyyah*	'moon-faeces'
Toba	*huaqajñi l-'atec*	'star 3SG.POSS-excrement'
Hawaiian	*kūkae-lio*	'excrement/dung-horse'
San Mateo del Mar Huave	*aonts potwit*	'excrete black.vulture'

table 15: Languages with a metaphorical term for 'mushroom' on the basis of 'faeces'

A natural reaction to this data is to question whether the terms given in the consulted sources are really the generic terms for mushrooms as opposed to the name of a particular kind of mushrooms. However, this does not appear to be the case, since for instance the San Mateo del Mar Huave term is explicity glossed as "los hongos" suggesting a generic function, and in the other sources there is no indication either that a specific kind of mushroom is designated by the above terms.[9]

Moreover, there are four instances in which celestial phenomena are expressed by terms on the basis of 'faeces': 'meteoroid,' 'cloud,' 'horizon,' and 'dusk.' While at least the connection with 'meteoroid' is motivatable, the other associations are fairly unclear, at least at first glance. It is probably an areal phenomenon of New Guinea to have complex words for 'cloud' consisting of terms for 'wind' and 'faeces.' The other two remaining patterns are found in Austronesian languages. In Tetun, 'dusk' is *loro-teen* 'sun-excrement' (the term also denotes a species of moss), while in Lenakel, 'horizon' is alternatively *noua-nɨsii-tehe* 'fruit-excrement-sea' or *noua-nɨsii-neai* 'fruit-excrement-sky.' The Tetun data appear to be explainable by assuming that the last light of day seen at dusk is something the sun has left behind -excreted- before disappearing, and similarly, one could speculate

[9] In addition, there is some more, albeit not very compelling, evidence for a global prevalence of the association. Hladký (1986: 11) mentions that "the Czech word *houby* 'mushrooms' functions as an euphemism for *hovno* 'shit,'" although this fact is not necessarily sufficient evidence for a connection between the two meanings, since replacement of swearwords can also occur with phonologically similar words not standing in any semantic relation with them, as in Engl. *shoot* replacing *shit*. English may be another case in point, since *stool* as found in *toadstool* can, as also suggested by Hladký (1986: 15), mean inter alia "a discharge of faecal matter of a specified colour, consistency, etc.; the matter discharged" (Oxford English Dictionary), in other words, 'faeces.' However, the earliest attestation of *toadstool* in the Oxford English Dictionary (as *tadstoles*) dates to 1398, whereas the earliest attestation of *stool* in the sense cited above is from 1597 and occurs in a medical context, so that it is possible that when *toadstool* was coined, the sense of *stool* understood was indeed that of 'seat' rather than 'faeces.' At any rate, the underlying metaphor seems generally to be that mushrooms are the residua of the respective entities that form the second member of the compounds; also noteworthy in this context is the association with 'fart' in Kiliwa and perhaps in Yay (see Appendix E, 40).

that in Lenakel the horizon, as the perceived end of sky and sea, "is" its excrement. Generalizing, one can note the following metaphorical transfer patterms:

(i) From 'faeces' to other body secretions with a mostly negative connotation
(ii) From 'faeces' to secretions of animals ('wax,' 'honey')
(iii) From 'faeces' as the result of digestion to other natural processes which leave manifest remnants ('ash,' 'smoke,' etc.)
(iv) From 'faeces' to natural phenomena which, loosely, may be conceived as the excretion of some entity

In many cases in which a physical object in the broadest sense is expressed using this extension, it tends to be rather smallish in size and to have rather well-defined boundaries. This is true of the concepts 'calf of leg,' 'cloud,' and 'mushroom.'

Another noteworthy point is that often more than one instance of the same conceptualization strategy is found in the same language. This points to the possibility that the distribution of lexical patterns such as this one is not fortuitous. More generally speaking, it points to the existence of abstract underlying schemes of semantic processes in word-formation present in one language, but not in another (see also § 6.2.3.1. on 'eye'). This is in principle reconcilable with cognitive accounts of word-formation as outlined e.g. by Tuggy (1987, 2005), albeit in a quite different manner. While such cognitive approaches to word-formation highlight the abstract nature of schemes on the formal side, here there appears to be a case of an abstract semantic pattern that can be realized in a number of ways on the semantic side.

Lexical Item	Gloss	Meaning
aonts xiül	'excrete tree'	'resin'
aonts mijiw-aran	'excrete breast'	'milk'
aonts potwit	'excrete black.vulture'	'mushroom'
mi-xiüt aonts	'AL.POSS-line excrete'	'guts'
aonts najloc	'excrete wound'	'pus'
aonts ombeayaran	'excrete mouth'	'saliva'
aonts oxingueran	'excrete nose'	'snot'

table 16: complex lexical items in San Mateo del Mar Huave involving *aonts* 'excrete'

To illustrate what is meant, it is instructive to look at a number of complex lexemes in San Mateo del Mar Huave (table 16), which all contain *aonts* 'excrete' (one of the San Mateo del Mar Huave terms for 'faeces,' *aonts-aran*, is based on this root, *-aran* being a suffix indicating (probably inalienable) possession, compare Stairs Kreger and de Stairs 1981: 291). While there is frequent parallelism in form (but note the differing structure of the word for 'guts'), even more striking is that the same pattern of semantic transfer is employed to conceptualize a wide variety of disparate referents.

6.2.3.4. Kinship semantics and their extensions

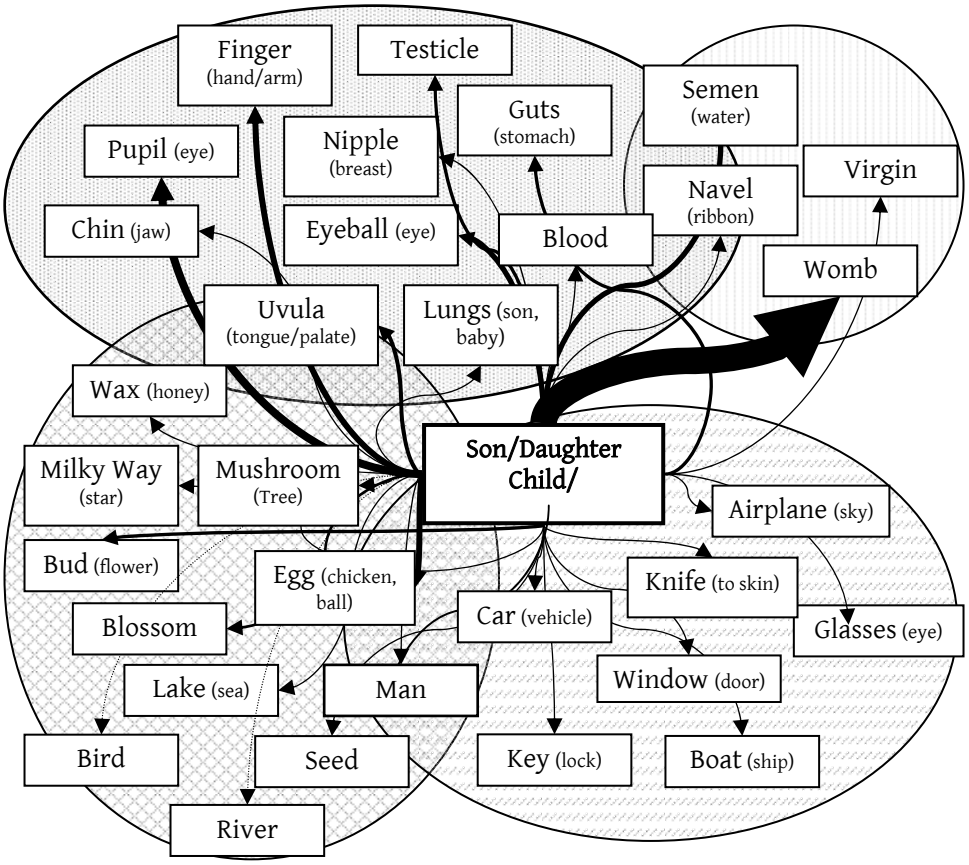


fig. 12.: lexico-semantic associations for 'son,' 'daughter,' or 'child'

The diagram in figure 12 provides an overview of colexifying and morphologically complex terms in which either of the meanings 'son,' 'daughter,' or 'child' generally is expressed by one constituent, with the meaning of (one of the) other term(s) in complex terms given in smaller size in parentheses.

As in the diagrams in §§ 6.2.3.1. – 6.2.3.3., contiguity-based associations are in the upper right corner. In this case, there are such associations with four of the meanings on the meaning list: some terms for 'virgin' contain an element meaning 'girl,' and frequently, terms for the 'womb' are complex featuring elements with either of the meanings 'son,' 'daughter,' or 'child' (see Appendix E, 148). Furthermore, in Miskito, *won klua* 'navel' is analyzable as 'child ribbon' (and probably also denotes the 'umbilical cord,' although it is not explicitly glossed so), and in Mbum, 'semen' is *mbìì gûn* 'water child' (Kiowa colexifies 'semen' with 'child,' with the optionally complex term *'iH-t̂Hę* 'child/semen/egg-white' for 'semen'). The rest of the cross-linguistic associations can be classified as being metaphorical in nature. For the majority of associations, the general underlying transfer pat-

tern can be, inspired by Jurafsky's (1996) analysis of semantic sources for diminutives, summarized as CHILDREN ARE SMALL THINGS. Thus, one finds denominations for smallish body parts such as the 'finger' like Katcha *bibala ma nizo* 'child GEN hand,' for the 'uvula' like Tetun *nanarak-oan* 'palate-son,' and so on (the association with 'lungs' is only found in Dadibi, where *ogwa wai* 'spirit of man, lung' appears to be analyzable as 'son baby'). The associations with 'pupil' and 'eyeball,' while arguably also metaphorical in nature, are likely based on a slightly different perceptual property of the pupil, namely to reflect a small image of oneself in one's interlocutor's eye (Tagliavini 1949, Brown and Witkowski 1981, Urban forthcoming) that yields terms for the 'pupil' on the basis of meanings such as 'child,' 'small person,' 'doll,' etc. As far as the association with 'child' specifically is concerned, however, it can be reconciled with the general association of 'child' ('son,' 'daughter') with small things. As with body-parts, CHILDREN ARE SMALL THINGS arguably also underlies complex terms for nature-related terms on the basis of 'child,' like those for 'mushroom,' 'bud,' 'egg,' 'lake,' and 'Milky Way.' It is interesting to note that, with the exception of the 'Milky Way,' in effect an agglomeration of distant and hence small stars, these referents are roundish in nature, and this seems to be nonaccidental. Matisoff (1992: 304), in discussing complex terms in Thai based on *lûuk* 'child' (including for instance *lûuk-faj* 'child-fire' = 'spark'), maintains that the semantic development of *lûuk* was from 'child' first to 'fruit' (with the additional conceptual similarity that fruits serve the reproduction of plants, just as children do in the case of humans) to 'small thing' in general.

More difficult to analyze is the association with 'wax' (Kanuri *kàmàgàn-mí* 'honey-son.of'). Perhaps this is because 'honey' is more desirable than 'wax'? Likewise, in the domain of artifacts, 'keys' are smaller than the 'locks' they are used to open and close, a 'window' is smaller (but similar) to a 'door,' a 'boat' is smaller than a 'ship' (but used for the same purpose), a 'car' may be conceived of as a small 'vehicle' when the standard of comparison is, say, 'trucks,' and an 'airplane' can be seen as a small point on the sky (Ngaanyatjarra *yilkaringkatja* contains *yilkari* 'air' and *katja* 'son'). The same language is responsible for the association with 'glasses' by the term *kurungkatja* (*kuru* 'eye,' *katja* 'son'). The association with 'knife' is due to an advanced stage of conventionalization in Kiliwa, where *na(y)* in fact colexifies 'child' and 'small;' the relevant term is analyzable as *na(y)-c-ruuw* 'child/small-INST/MOUTH-to.skin' and the "literal" translation offered in the consulted source is "small skinner"). In a few cases, however, the direction of the mapping is apparently not from 'child' (or 'son, daughter') to the other meaning, but rather the other way around: for instance, in Samoan, *gāʻau* 'guts, intestines' is also a jocular designation for 'son.'

A diagram showing the associations found in the sample for 'mother,' 'father,' or 'parents' generally is in figure 13.

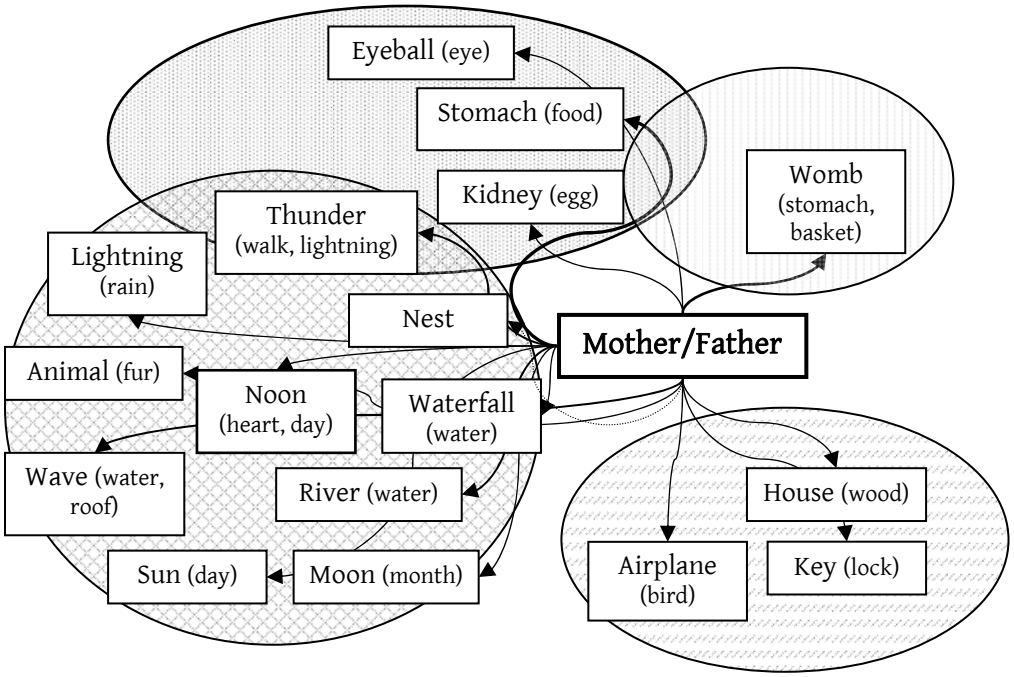


fig. 13.: lexico-semantic association for 'mother' and 'father'

Contiguity-based associations realized by complex terms are restricted to precisely one meaning, which is, as in the case for 'child,' the 'womb': Ket has *ām-d huūj* 'mother-POSS stomach' and Bislama *basket blong mama* 'basket GEN mother/pregnant.'

Jurafsky (1996: 546) proposes, among others, the metaphorical transfer pattern BIG THINGS ARE MOTHERS, which would account for the complex terms for 'eyeball,' 'river' (note that frequently, terms for 'river' contain morphemes meaning 'big' and 'water,' see Appendix E, 47), and 'house,' which is a large structure made of 'wood,' and presumably also 'wave' and 'waterfall.' Another pattern, perhaps related to Jurafsky's (1996: 547) GROUPS ARE FAMILIES, in which the opposition between 'mother' = 'big' and 'child' = 'small' plays as central role, is the extension of 'mother' to parts of an object situated in the center of that object, which would account for the associations with 'eyeball' and 'noon.' The association with 'lightning' is due to Mbum *máà-mbàm* 'mother-rain,' and those with 'thunder,' 'sun,' and 'month' are found in one and the same language, San Mateo del Mar Huave: *ajüy teat monteoc* 'walk father thunderbolt' is the term for 'thunder,' *müm caaw* 'mother month' that for 'moon' and *teat nüt* 'father day' that for 'sun.'

It must be noted that there are a number of cases where presence of 'mother' or 'father'-terms remains conceptually somewhat unclear. Koyraboro Senni has *duma-ñaa-guuri* 'kidney-mother-egg' for 'kidney,' and Anggor has *nɨnɨhondɨ* 'animal, game animal,'

presumably analyzable as /nɨne-hondɨ/ 'fur/feather-mother.'[10] Also puzzling is Manange *1tʌntsʌ-1amʌ* 'lock-mother' for 'key'[11] in the light of Matisoff's (1992: 300, 306) examples from languages of Southeast Asia and Oceania where the mother-child opposition is exploited to express the meanings 'lock' and 'key,' such as Thai *mêe-kuncɛɛ* 'lock, padlock' (*mêe* 'mother') – *lûuk-kuncɛɛ* 'key' (*lûuk* 'child') and Malay *ibu kuntji* 'master/skeleton key, lock' (*ibu* 'mother') – *anak kuntji* 'key' (*anak* 'child'). In fact, these languages are not the only ones in which such an opposition is found, and the existence of such patterns has not gone unnoticed by both Matisoff (1992) and Jurafsky (1996). 'Thumb' is 'mother of hand' and the 'finger' 'child of hand' "throughout the Mayan family." For instance, Tzeltal has *smeʔak'ab'* 'your thumb' (literally "its mother your hand") and *yalak'ab'* 'your finger' ("its child your hand," Matisoff 1992: 346fn97, quoting Terrence Kaufman p.c.). This is confirmed by the sample data: In Itzaj, *al k'ab'* 'finger' is analyzable as 'child arm/hand' ('thumb' is not on the meaning list). Again in Malay, *ibu panah* is 'bow' and *anak panah* 'arrow' (Matisoff 1992: 301), and in Meyah, *otkonú efesá* 'stomach child' is 'intestine,' and *otkonú mosú* 'stomach mother' is 'large intestine.'

Curiously, it is much more frequently the female parent, the 'mother,' rather than the 'father,' whose designant is used in complex terms to convey size or importance. This difference also turns up in a different context, namely that evidence from several languages suggests that if the opposition is not 'mother' – 'child,' but 'female' – 'male' (these competing patterns have also been noted by Matisoff 1992 and Jurafsky 1996), it is still the female semantics associated with big size and the male semantics with normal or small size.[12] In Mali, the masculine and feminine noun classes are employed in a similar fashion, with the feminine class conveying big size and the masculine normal size: thus *churet-ki* (*-ki* is the feminine noun class marker) is 'a large flame' and the masculine *churet-ka* 'an average sized flame' (Stebbins 2005: 101); for particularly small objects the diminutive class suffix is *-ini* is chosen in Mali (that indeed gender is a factor in the semantics of the noun classes is shown by examples such as *asingal-ka* 'male forest spirit' and *asingal-ki* 'female forest spirit,' Stebbins 2005: 103, where size difference does not seem to play a role). In Bora, Manguaré drums used for communication come in a set of a big and a small one, the big one being called 'female' and the small one 'male' (Seifart and Meyer 2010: 4). In Yeli Dnye, 'hisfingers' are *kóó pyââ dmi* 'hand/arm woman bundle' (where 'bundle' is a classifier) and 'his thumb' *kóó kːaa pyââ* 'arm taro woman' (Levinson 2006b). In Koyraboro

[10] Note also Samoan *tamatama'ilima* 'finger,' containing *lima* 'hand' and *tama'i* 'small thing.' *Tamā* is 'father,' but the initial two syllables of the complex term seem more likely due to partial reduplication of *tama'i*.

[11] For 'lock' Manange also has a redundant compound *1tʌntsʌ-2tsʌtsʌ* 'lock-young' which again suggests the opposite direction from the apparently more widespread Southeast Asian pattern by virtue of featuring *2tsʌtsʌ* 'young' (*1tʌntsʌ* itself is perhaps borrowed from Nepali *taalcaa* 'lock').

[12] In Khoekhoe, different nominal designants indicating gender and number may be suffixed to one and the same root to yield different meanings. Here, the evidence from the present data is somewhat inconclusive, and gender assignment seems to be to a large part arbitrary, but there is slight evidence that it is indeed the feminine gender associated with smallness or lesser "fierceness:" *llnâ-b* (*-b* being the nominal designant for the third person singular masculine) is 'horn of an animal', *llnâ-s* (*-s* indicating third person singular feminine) is 'trumpet, brass instrument,' *lnanu-s* is 'rain, raincloud,' *lnanu-b* 'rain, thunderstorm,' *ǂkhara-s* is 'testicle,' *ǂkhara-b* 'scrotum.'

Senni, the 'mother' – 'child' and the 'man' – 'woman' opposition coexist. Compounds with *-ñaa* 'mother' denote "a complete object that constitutes the source of the entity denoted by the compound initial" (Heath 1999: 108, compare also Matisoff 1992 on the extensions of 'mother' to 'source, origin'). Thus, *dugu* is 'incense' and *dugu-ñaa* is the 'incense plant;' plants and their products are the most common application of the system. In contrast, compounds with *-ize* 'child' have diminutive semantics (*tasa-yze* 'small bowl') or fulfill the 'unit-excerpting' function of the diminutive noted by Jurafsky (1996) in the realm of botany, as in *hayni-ize* 'grain of millet,' but also in body-part terms, as in *himbiri-ize* 'single hair' versus *himbiri* 'hair' collectively and *kabe-ize* 'finger' as opposed to *kabe* 'hand.' Here, corresponding pairs as in other languages as discussed above are found: *daarey-yze* is the 'jujube fruit,' and *daarey-ñaa* the 'jujube tree' (Heath 1999: 108). There are also compounds with *aru* 'man' and *woy* 'woman,' which are used to specify gender of an animal (*čirow* 'bird,' *čirow-aru* 'male bird,' *čirow-woy* 'female bird'), but also to distinguish size: *hoŋko* 'water lily fruit,' *hoŋko-garu* 'large water lily fruit,' *hoŋko-woy* 'small water lily fruit' (Heath 1999: 109). Thus, in Koyraboro Senni, it is the compound with 'man' that conveys bigness and that with 'woman' conveys smallness, in contrast to the evidence from languages discussed above. This is the paradoxical situation noted by Jurafsky (1996: 545) that 'woman' or more generally 'female' may be associated cross-linguistically with big as well as small size. The evidence for this presented by Jurafsky comes mainly from gender alternation, such as Hindi *ghantā* 'bell' (masculine) and *ghantī* 'small bell' (feminine) rather than from morphologically complex terms. Such instances are not found in the present sample, and the evidence from the sample suggests that the extension of 'woman' to big size is more common than that with small size, especially as far as complex terms of the lexical type are concerned,

There is also one instance where a term for 'grandmother' rather than 'mother' is used to convey big size: in Mali, 'flood' is *milat-ka av-uouk* 'coconut.shell-M.SG SG.POSS-grandmother;' furthermore, the Rama terms *dama árkali* 'lightning' and *dama yatangi* 'thunder' contain *dama* 'grandfather.'

6.3. ENVIRONMENTAL FACTORS

Of course, especially in the conceptualization of the natural surroundings by a speech community, the properties of these may play a role in the development of certain patterns of colexification. One example is the colexification of 'forest' and 'mountain' which is attested in a number of languages in the sample (unsurprisingly, semantic shift in diachrony is also attested, for instance in Uralic, Redéi 1988: 571). There is a straightforward explanation for this pattern available: in mountainous terrain, dense vegetation with larger plants is found on mountain slopes, whereas in the valleys, vegetation is more sparse, typically grassland (compare the Kiliwa term *?+mat=xu?sawy* 'DN+earth/land=clean/clear' for 'valley'). Obviously, such a landscape is likely to trigger the colexification of 'forest' and 'mountain' (see also Fränkel 1938 for semantic shift between terms for environmental features within Indo-European). It is then a fine example of spatial contiguity semantically – where there are mountains there is forest and vice versa.

Another lexical feature that may be directly due to environmental features is the colexification of 'cloud' and 'fog.' Laycock (1970: 1138) states that in New Guinea, this pattern occurs typically in high mountain areas. It is empirically testable whether there is indeed a correlation between this pattern of colexification and altitude on a global scale. Data for altitude were gathered for this purpose from GTOPO30, a digital elevation model of the world available at http://www.gpsvisualizer.com/elevation.html (latitude and longitude data from the World Atlas of Languages Structures were used for the altitude query).[13] Under the interpretation that there is a scale of lexical differentiation between the two referents, from fully differentiated when lexically unrelated terms are present via semi-differentiated when a complex term for 'fog' is present on the basis of 'cloud' (as described in Appendix E, 25) to lack of lexical differentiation in the case of colexification, there is a statistically significant correlation between altitude and the degree of differentiation in the languages of the statistics sample (S = 33422.35, Spearman's ρ = .3894346, p = .0009417), but even when one removes analyzable terms from the calculation due to their ambiguous status, there still is a strong difference, significant at p = .00108 (S = 26237.3, Spearman's ρ = .3993293), to the effect that languages spoken at higher altitudes are more likely to feature this pattern. A visualization of the differences is provided in figure 14.

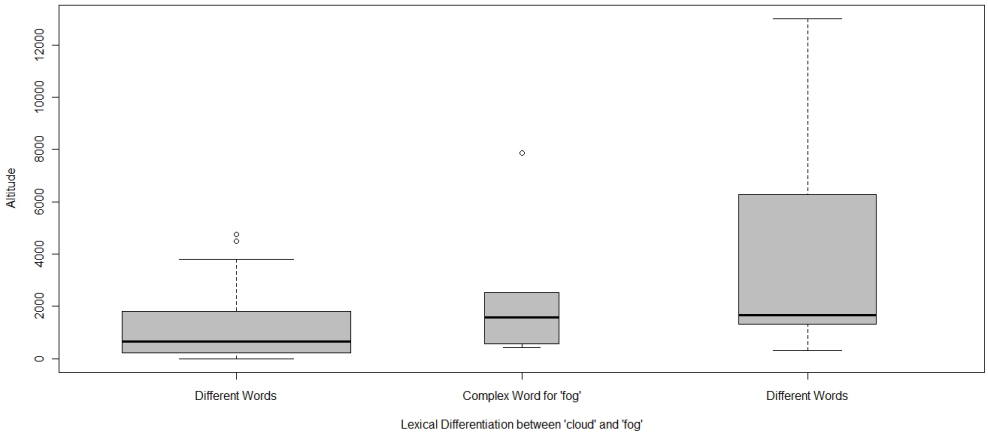


fig. 14: colexification of 'cloud' and 'fog' depending on altitude.

However, there are also a lot of languages spoken at very high altitudes which use different words for the two referents, and the statistics is not sensitive to alternative synonyms or near-synonyms that may be present in languages with colexifying terms.[14]

[13] Unfortunately, the query for unknown reasons yielded errors for the coordinates for two of the languages, Embera and Bislama, which is why they are excluded from calculation.

[14] Another factor was tested, but with negative results: One can also speculate whether there are extra-linguistic factors that might predict a language's behavior with respect to the carving up of the lexical domain of bodies of

6.4. GENEALOGICAL AND AREAL LEXICO-SEMANTIC PATTERNS

6.4.1. INTRODUCTION

This section addresses another major topic concerning the lexicon from a cross-linguistic point of view. Alongside asking questions about the distribution of quantitative aspects of lexical motivation that were discussed at length in chapter five, it is at least equally interesting to ask about the distribution of individual lexico-semantic patterns, both from a genealogical and areal perspective.

6.4.2. INTRA-FAMILY COMPARISONS FOR SEMANTIC ASSOCIATIONS

From the point of view of genealogical linguistics, an interesting question that can be asked is whether there are lexico-semantic patterns that are peculiar to a particular language family. Departing from this basic question, it is possible to extend the discussion into more theoretical matters, that is, to elucidate whether patterns in semantics and in the structure of morphologically complex lexical items can be diachronically stable within language families so as to make them useful additional features that can, alongside e.g. regular correspondences in phonology, be of use for genealogical classification and the establishment of language families. Laycock (1975: 228), for instance, summarizes a number of areal patterns of New Guinea reported in Laycock (1970), and notes that "[i]t seems that some of these distributions may be highly regional, and may prove, if charted on the language map of the New Guinea area, to be useful in the establishment of linguistic supgroupings" and goes on to say that "[s]uch a use of semantic domains for linguistic taxonomy is a new approach which is as yet untried, but which shows signs of promise for the future." Matisoff (1978: 231) suggests that "[t]he shifting patterns of semantic association within a language or a language-family are at least as interesting as phonological changes through time, and may prove to be equally criterial for establishing degrees of genetic relationship" (see also François 2010 on the reconstruction of semantic patterns and the issue of disentangling them from areal diffusion).

Within-family stability of lexico-semantic associations were investigated selectively for two well-established and uncontroversial families, namely Tupian (actually,

---

water, in particular with respect to the colexification of 'water' and 'river.' It is for instance possible to speculate that in languages with comparably many speakers which are thus presumably spoken in a relatively larger territory when compared with languages with a small speech community, people will be familiar with several different rivers in the territory, whereas in a smaller language in terms of speaker size, confined to a small territory, there may well only be one natural watercourse people encounter on a day-to-day basis, serving as the source of fresh water. To assess whether empirical data support this hypothesis, figures for the size of the respective speech communities were gathered from Lewis (2009), see appendix D for data. However, statistical analysis does not yield a significant difference between the populations under the interpretation, as in the analysis for 'fog' and 'cloud' above, as a continuum of decreasing lexical differentation ($S = 54230.14$, $\rho = .05117413$, $p = .674$, Wilcoxon rank sum test). Probably the initial hypothesis is too simplicistic anyway, given that small speech communities need not be settled in a particular territory, but instead may be highly mobile (Johnson and Earle 2000).

Tupi-Guaraní more narrowly) and Uto-Aztecan. For this task, a quite simply methodology was employed: when for a given meaning equivalents are available for all sampled languages within the language family, and where at least one has a morphologically complex expression or particular pattern of colexification, the other language(s) was/were checked for presence or absence of the relevant pattern.

The Tupi-Guaraní languages Guaraní and Wayampi have a total of 24 patterns in common, most of which are found in the domain of nature-related and topological terms. They include the following pairs of colexification: 'nose' – 'beak' – 'prow of canoe,' 'bark' – 'skin,' 'coal' – 'embers,' 'foam' – 'bubbles,' 'grass' – 'plant,' 'river' – 'water,' 'smoke' – 'steam,' 'tree' – 'wood,' 'boat' – 'canoe,' 'paper' – 'book,' 'nipple' – 'breast, teat,' 'fingernail' – 'claw' and extension of 'skin' to 'surface, cover' more generally. As for complex terms with common structure, there are complex terms for 'cave' ('stone-hole'), 'Milky Way' ('tapir-way'), 'star' ('moon-fire'), a term for 'belt' containing the word for 'waist,' complex terms for 'tear' ('eye-water'), words for 'whirlpool' containing verbs meaning 'to twist' and 'to turn around,' and complex terms for 'milk' with constituents 'breast' and 'water, liquid, juice.' Moreover, there is an association between 'horn' and 'point' by colexification in Guaraní and by a complex term in Wayampi, and of 'semen' with 'child' by colexification in Wayampi and by a derived term in Guaraní.

Many of the shared patterns are so frequent cross-linguistically that their value as a characteristic of a genealogical grouping is strongly diminished as they might easily also have come into being independently (see relevant sections in Appendix E). However, in Tupi-Guaraní there is also a genealogical signal consisting of rare or even absent structures outside of this family, in particular the terms for 'Milky Way' and 'star.' Based on glottochronological calculations, Rodrigues (1964) dates the split of Tupi-Guaraní to approximately 2,500 BP, and Silva Noelli (2008: 663) informs that radiocarbon dating of artifacts from archaeological sites even suggests a "much earlier" date. Another unrelated piece of evidence for an at least relative stability of certain lexico-semantic associations comes from Malagasy. The Malagasy word for 'sun' is morphologically complex and is of the type 'eye of day,' which is typical for languages of the Austronesian family and languages of Southeast Asia (Urban 2010). However, Malagasy has replaced its inherited word for 'eye' by a loanword from Bantu, *màso*, but retained notably enough the complex term for 'sun' of the Austronesian-Southeast Asian type employing the Bantu loanword. This shows that lexico-semantic structures at least have some potential for diachronic stability over the not inconsiderable time-depth of probably more than one millennium in the case of Malagasy 'sun' (Adelaar 1989: 35 tentatively posits a migration of Austronesian speakers to Madagascar in or after the 7th century AD, while noting that there is also evidence for prolonged contacts with South Sumatra after the migration event) and at least two and a half millennia in the case of Tupi-Guaraní.

However, there is also a lot of negative evidence. The Uto-Aztecan language family has considerable greater internal diversity than Tupi-Guaraní, and its common ancestor is thought to have been spoken earlier than that of Tupí-Guarani. Holman et al. (2011) estimate a breakup date around 4,000 BP. The evidence from this family is sobering, in spite of the high similarity in the abstract values (percentage of analyzable terms and of

metaphor-driven terms) for this particular family: there is not a single pattern that is common to all of the four languages in the sample, and the four cases of a correspondence of three languages, colexification of 'skin' and 'bark' in Comanche, Pipil, and Yaqui, 'sun' and 'day' in Cahuilla, Comanche, and Yaqui, 'tree' and 'wood' in Cahuilla, Comanche, and Pipil, and of 'moon' and 'month' in Cahuilla, Comanche, and Yaqui, are all among the most common semantic associations in the world's languages as a whole. Even correspondences between two languages are not very common: Patterns of colexification shared by two of the four sampled Uto-Aztecan languages include: 'gold' – 'money' (Comanche, Pipil), 'water' – 'river' (Cahuilla, Pipil), 'pit' or 'seed' – 'eye, face' (Cahuilla, Pipil, in which latter the meaning 'face' is restricted to compounds), 'land' – 'earth' (Pipil, Yaqui), 'boat' – 'canoe' (Comanche, Yaqui), 'fingernail' – 'claw' (Pipil, Yaqui), 'skin' – 'leather' (Yaqui, Pipil), and 'skin' – 'shell' (Pipil, Comanche). As for complex terms, Cahuilla and Pipil have a complex term 'mouth-hair' for 'beard,' and there are complex terms for 'forest' with a constituent meaning 'tree' in Comanche and Pipil, which are however structurally quite different otherwise. Also, Comanche and Pipil both have terms for 'sky' based on the notions 'high' or 'above,' and there are complex terms for 'scissors' containing a verb meaning 'to cut' in Comanche and Yaqui. Further more heterogeneous commonalities include an association between 'honey' and 'candy, sweets' by colexification in Yaqui and by a complex term in Pipil, and a complex term for 'lake' involving a constituent meaning 'water' in Cahuilla, and semianalyzable terms of that kind in Comanche and Yaqui. Mirroring the colexification of 'water' and 'river' in Cahuilla and Pipil, Yaqui has the complex term batwe /ba'a-bwe'u/ 'water-big.' There are complex terms for 'swamp' with one constituent meaning 'water' in Pipil and Yaqui, but with otherwise different structure (and there is a semianalyzable term in Comanche). Comanche colexifies 'mouth' and 'lip,' and a complex term for 'lip' with a constituent 'mouth' is featured in Yaqui. Finally, there are complex terms for 'nostrils' with a constituent 'nose' in Pipil and Yaqui, and complex terms for 'eyebrow,' 'eyelash,' and 'eyelid' with one constituent meaning 'eye' in many of the languages but with varying other constituents. Very many of the abovementioned associations are common cross-linguistically and well attested outside Uto-Aztecan (see relevant sections in Appendix E), another one, the association of 'sky' with meaning like 'high, above,' is common in North America as a whole, and still others, such as colexification of 'gold' with 'money' and the terms for the 'scissors' cannot be interpreted reasonably as indicating a deep historical signal because they almost certainly postdate the time of European contact.

As a preliminary conclusion from the small set of investigated families, the genealogical signal of lexico-semantic structures is weak, and thus appears to be of rather limited use for traditional historical linguistics concerned with single language families alone (albeit not of no use at all, as the case of Tupi-Guaraní shows, though in general this family "is noted for a high degree of lexical and morphological similarity among its member languages in spite of their extensive geographical separation," Jensen 1999: 128). However, this does not mean that lexico-semantic patterns are not amenable at all to historical interpretation (note also the areal skewing with regard to lexical differentiation of 'sun' and 'moon' which is amenable to a historical interpretation). This is one of the topics to be

discussed in the following section, which deals with areal factors that are at times responsible for the distribution of individual patterns.

6.4.3. AREAL PATTERNS IN SEMANTIC ASSOCIATIONS

6.4.3.1. Introduction

Alongside intra-family comparison, on the other hand, one can also take the opposite perspective and ask whether certain patterns in the lexicon are not genealogically but still areally restricted. If such patterns are found, this would be an indication that the lexicon is susceptible to significant influence from neighboring languages in language contact situations. Evans (1990: 137), for instance, maintains "that the whole continent of Australia is characterisable as a linguistic area from the point of view of certain types of polysemy and semantic change that are common right across the continent, but rare or unreported elsewhere." As the quote already makes clear, areal influences in the lexicon pertain both to morphologically complex lexical items as well as colexification, although the latter has not received ample discussion: when it comes to influences languages in contact can exert on each other with respect to the lexicon, what immediately comes to mind is lexical borrowing, that is the transfer of a word, in its phonological form and typically also semantic content, from one language to another. What probably comes to mind next are calques or loan-translations, and indeed, there are a number of studies devoted to areal calquing exclusively or as part of larger discussions of linguistic areas (to be mentioned in a minute). In addition, there is an ugly duckling that has not received the same amount of attention: this is the transfer of semantic structure from one language to one or more other neighboring languages and the convergence of neighboring languages with respect to the internal semantic structure of their vocabulary items. In contrast to lexical borrowing, this process does not involve transfer of linguistic material, but rather of semantic structure alone, which is then superimposed onto native lexical items. Saying that this phenomenon has not received wide attention does not mean that the phenomenon is not known in theory, but discussions in the literature are typically not longer than half a page. In the German literature, "Lehnbedeutung" is typically used (e.g. Blank 1997: 349); this term probably goes back to Betz (1949). Haugen (1950: 219) uses the term "loan synonym" for this process "which only adds a new shade of meaning to the native morpheme;" Curnow (2001: 427), in a brief discussion, adopts Haugen's terminology, and defines loan synonymy as the process "where the meaning of a word is extended to fit the pattern of lexical extensions of a word in another language with a similar basic meaning." Geeraerts (2010: 29) speaks of "semantic borrowing," "the process by means of which a word x in language A that translates the primary meaning of word y in language B copies a secondary meaning of y," noting that the process is also known as a "semantic calque." One example adduced by Geerarets (2010: 29-30) is Greek *angelos* which acquired, under the influence of (translation of religious texts from) Hebrew *ml'k*, the additional meaning 'heavenly messenger,' i.e. 'angel' alongside its basic and original meaning 'human messenger, envoy.' Smith-Stark (1994: 17) uses "loan shift" to refer to "calquing the internal structure of a lexical item." Another major relevant line of research is that concerning relexification in the process of creole genesis, one of the facets of which is that the internal lexical seman-

tic structure of lexical items from the substrate language is mapped onto the phonological shape of words of the lexifier language (for an overview see e.g. in Lefebvre 2001 and Lefebvre 1998 for a more detailed case study). Relexification is primarily adduced to account for properties of creole languages.

Also related but slightly different because concerned with polysemy of grammatical morphemes in grammaticalization processes is Heine and Kuteva's (2005: 100) notion of "polysemy copying," as well as Mous's (2003) "lexical manipulation" in language mixing.

The terminological multiplicity reflects the paucity of detailed influential research on this phenomenon. Empirically oriented discussions of semantic transfer are mostly equally short as their theoretical counterparts, with the exception of Enfield (2003), who offers a detailed case study of semantic and grammatical convergence in Mainland Southeast Asia. Moreover, Ross (1996) mentions the remodeling of the lexical semantics of Takia vocabulary under influence of Waskia in the process of the "metatypy" of Takia; this includes loan translation, but Ross also explicitly mentions semantic transfer. Similarly, Aikhenvald (1999: 406) offers a brief discussion of lexical semantic influence of East Tucano languages on the Arawakan language Tariana, and recognition of the phenomenon is implicit in Campbell et al. (1986) and Matisoff (1978, 2004).

The following discussion of areal lexico-semantic patterns is concerned both with loan translations as well as with semantic transfer, in short, any lexico-semantic patterns comprised under the definition of lexical motivation as described in § 3.5. Matisoff (2004) uses "areal semantics" to refer to both sorts of patterns; however, strictly speaking this is a misnomer, since loan translations in fact are not actually concerned with semantics proper (compare Marty's 1908 critique of the equation of semantic patterns within complex expressions with semantics itself mentioned in § 2.4.), which is the reason for speaking about areal lexico-semantic patterns rather than areal semantics as a cover term. Where lexical semantics belongs in the general theory of language contact is likewise unclear. Thomason and Kaufman (1988: 121), unsurprisingly given the sparseness of dedicated literature on the topic, report not having found many examples of contact-induced change in lexical semantics, but state that they would expect it to pattern with changes in phonology and syntax, which they show to occur even if contact is not very intense. However, the motivation for convergence in the lexicon seems to be the same as for contact phenomena in general: it is an adaptation serving to increase intertranslatability of the languages in the repertoire of bi- or multilingual speakers (Gumperz and Wilson 1971), and, as Sasse (1985: 84-85) points out, lexical and grammatical calquing and syntactic convergence share this same ultimate cause.

While this is not the first study on this subject, it appears to be the first one with global scope. Past research on areal patterns in the lexicon has at times suffered from a lack of informedness on the cross-linguistic distribution of the phenomena in question. As Evans (1990: 152) also notes, claims concerning areality such as the one made by him for Australia as a whole "can only be fully substantiated when proposed paths of semantic change and grammaticalisation in Australia can be shown to differ significantly from those found in other language families and areas." For instance, Campbell (1979) as quoted in Smith-Stark (1994: 18-19), mentions 'skin-tree' for 'bark' and lack of a separate word for

'river' as areal patterns of Mesoamerica. The data of this study show, quite to the contrary, that these phenomena are widespread globally (Campbell also mentions other patterns that may truly be characteristic of Mesoamerica in particular). Campbell et. al (1986: 553, table 1) provide a list of 55 lexico-semantic patterns that are candidates of being diagnostic of the Mesoamerican linguistic area, of which twelve survive closer scrutiny (that is, examining whether a given pattern occurs in control languages immediately to the north and the south that are not part of the putative area) and are accepted by Campbell et al. as areal features of Mesoamerica. Among them are 'vein' = 'road (of blood),' 'edge' = 'mouth,' and 'lime' = 'ash,' 'stone-ash.' As Campbell et al. (1986) themselves note, it can be imagined that the former two patterns might also be coined independently and remark in footnotes that they in fact are also found in South America. The data of the present study confirm this, and reveals further cases elsewhere. Similarly, Matisoff (2004), while correctly pointing out that certain associations (such as 'fire' + 'tongue' = 'flame') are common cross-linguistically because they "are so 'natural' to human though processes" (2004: 351), reports some lexico-semantic structures considered by him to be peculiar to Southeast Asia in the absence of known data from other areas. Among them are terms for 'meteor' on the basis of 'star' and 'faeces' in Hmong and Lahu (2004: 367). The present sample identifies this pattern in Sedang in Southeast Asia, but also shows that the association is also, and apparently more frequently, found in the Americas, namely in Central Yup'ik, Haida, Highland Chontal, and Toba as well as in the Austronesian language Tetun (see Appendix E, 36). Matisoff (2004: 367) also mentions a metaphorical transfer from 'eye' to 'anklebone' under the heading "Southeast Asian lexico-semantic areal features," but this association is even more frequent cross-linguistically than that between 'star' and 'faeces,' and occurs in very many areas of the world (see Appendix E, 99).

Here, the question is when one is willing to accept a particular feature, be it grammatical or lexical, as areal: is it enough if it is absent in the immediate vicinity of the area that is to be demonstrated, while it is acceptable if it shows up further South or North on the same continent or with some frequency in completely other areas of the world? Arguably, the above evidence does not mean that the respective patterns suggested by Matisoff (2004) should not be considered areal features of Southeast Asia, but they raise the question where one should draw the boundary at which one cannot speak of areal features anymore.

The assorted examples mentioned above point to two related vexing problems: the first concerns the need to test putative areal patterns against solid cross-linguistic evidence, or at the very least against a control sample of neighboring languages. The present data make it possible to assess areal lexico-semantic patterns against solid cross-linguistic evidence; and they also make it possible to offer for the first time a more extensive study of semantic transfers without concomitant transfer of linguistic material. The second issue already hinted to above is the problem of defining linguistic areas in the first place (for relevant discussion, see e.g. Bisang 2006, Muysken 2008, Campbell 2009). One big issue in that task is the problem of circularity: linguistic areas, by definition, are areas in which a number of languages have come to share structural features that are not due to genealogical inheritance, and the commonalities between these same languages is then

explained by their participation in the linguistic area. Furthermore, mere eyeballing of maps showing the distribution of a linguistic feature for potentially interesting areal patterns is methodologically dubious (Cysouw 2005, inter alia).

Because of the intricate difficulties in delimitating linguistic areas in a methodologically principled way, Bickel and Nichols (2009, in press) have taken a relatively radical measure: instead of relying on linguistic evidence in setting up areas, they first define areas without actually recurring to linguistic evidence. In this approach, geographically and sociohistorically delimited areas are first seen as hypothetical linguistic areas as well, and statistical procedures are then employed to test whether there is a linguistic correlate between geographically contiguous regions of the world which are known to be sociohistorically linked to each other.[15]

Here, a geography-based approach is adopted as well, as described in particular in Bickel and Nichols (2009: 487). Initially, a breakdown of the world in eleven areas modeled on that of Nichols (1992: 25-26) was used. These areas are: Africa (including Malagasy), Europe, Eurasia (excluding Europe, but unlike in Nichols 1992: 25-26, not the Caucasus), South and Southeast Asia (conventionally delimited, as in Bickel and Nichols 2009, by the Wallace Line), New Guinea, Oceania, Australia (including Tasmania), Western North America (including Kiliwa but excluding the Eskimo-Aleut language Central Yup'ik as in Nichols 1992: 25-26), Eastern North America (including Lesser Antillean Creole French and delimited from Western North America by the Rocky Mountains), Mesoamerica, and South America. This entails that the areas for which patterns can be detected are necessarily rather large, but since the density of the present sample is low, identification of comparably small sprachbund-sized areas is unfortunately not possible on the basis of the sample in the first place. For the same reason, the assessment of areality necessarily departs from classical definitions of linguistic areas such as that of Emeneau (1956: 15fn28) as quoted in Masica (2001: 209): "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the familes," which is not a feasible definition for present purposes since in many cases only one language per family was sampled in the first place. As is known today, linguistic areas come in all sizes, from very small to very large (as has emerged from studies such as Dryer 1989, 1992 and Nichols 1992), and it is of course possible, and even likely, that a given lexico-semantic association detected in the sample for only one language somewhere in the world in fact participates in a smaller linguistic area and might be a diagnostic feature of it. One instance would be the Vaupés area as described by Epps (2007), who also notes similarities in lexical structure. Another example of such a smaller linguistic area is the Clear Lake area in California, where speakers of Lake Miwok, Patwin, Wappo, and Pomoan languages lived in close proximity. Lake Miwok and Patwin belong to different subgroups of Penutian, while Wappo and Pomoan are unrelated or unrelatable by traditional methods both to one another as well as to Penutian. There were frequent con-

[15] A similar approach is sketched by Masica (2001: 219): "an alternative discovery procedure, ... and perhaps more 'objective', would be to start, not with languages or an area of interest, but with a few selected features of interest drawn from universal typology (i.e. rather than 'all' features), determine their distribution, and see what patterns emerge, particularly convergent patterns involving more than one feature" (emphasis removed).

tacts between the groups, intermarriage and widespread multilingualism (Callaghan 1964: 47fn3), the perfect medium for the emergence of linguistic areality. The Clear Lake area happens to be represented by Lake Miwok and Wappo in the present sample, and the Pomoan languages are represented by Kashaya, which is not immediately spoken around Clear Lake, but not far from it either. And indeed, there is evidence for convergence with respect to the lexicon (apart from a number of loanwords from neighboring languages in Lake Miwok and massive phonological convergence, Callaghan 1964). Some of the commonalities summarized in table 17 must be of fairly recent origin.

Concept	Lake Miwok	Wappo	Kashaya
'train'	*wikîi karéeta* 'fire wagon'	n/a	*hokare·ta* /ʔoho-kare·ta/ 'fire-wagon'
'eyeball'	n/a	*huči·-lél* 'eye-stone'	*huʔu· qʰaʔbe* /huʔuy qʰaʔbe/ 'eye rock'
'toilet'	n/a	*čéẏu čhùya* 'feces house'	*ʔahpʰahca* /ʔahpʰa-ahca/ 'shit-house'

table 17.: Some lexical commonalities in languages of the broader Clear Lake area

Returning to methodological questions, statistical evaluation of the evidence is not performed, the reason being that when the global statistics sample is broken down into smaller areas, the coverage of each area is very sparse, and the results one would arrive at in statistical analysis are not amenable to meaningful interpretation since the power of potential statistical tests would be extremely low. Furthermore, under these particular circumstances, statistical analysis would likely filter out a considerable number of genuine cases of areality as a direct consequence of the low sample density for each individual area. Instead, the following interim measure to assess areality is used: lexico-semantic associations are considered potentially areal if (i) the number of unrelated languages with the pattern in one of the areas outnumbers the number of unrelated languages in the entire rest of the world more than twice and (ii) at least 15 per cent of genealogically unrelated languages within the area in question feature the association (for the case of Australia, which in Dryer's 2005 classification hosts only one very large language family, Australian, only the second criterion is applied, since otherwise no areality could emerge for this continent by definition. Note also that the unity of Pama-Nyungan and non-Pama-Nyungan, and even the genealogical unity of Pama-Nyungan is not universally accepted, see Dixon 2002 and Evans 2005 as representatives of the different opinions and further references therein). The assessment of areality is carried out on the basis of the EXT-2 sample, since this task requires the densest possible coverage of the world's languages. As a measure of genealogical control, however, datapoints obtained from genealogically related languages are collapsed to 1. That is, genealogical inheritance, which apparently occurs at times (compare the Tupi-Guaraní words for 'star' and 'Milky Way'), is ruled out – an important requirement when areal rather than genealogical factors are at stake! This criterion is for consistency also applied to terms denoting artifacts, which are often unlikely to be of great age due to the recent introduction of their referents. The second criterion is a control for the size of the area in terms of distinct language families it hosts. Given that South America, for instance, has a much higher genealogical diversity than, say, Europe, it would be much easier for a particular association to be diagnosed as potentially

areal for that area than for ones with less linguistic diversity, and hence lower sample density. Without additional control there would be the danger that the number of diagnosed patterns is merely a function of genealogical diversity, which is not the case when the present criterion is applied.

The criteria have a certain probability of leading to a number of false positives, i.e. patterns that are found to be areal in spite of existing counter-examples reported in the literature, but not in the sample. For instance, one of the lexico-semantic patterns is the metaphorical extension of terms for 'beak' to 'prow of canoe' in languages of South America, next to one case in Hawaiian. As a comparison with the data for the association 'beak' – 'nose' quickly shows, usually the very same terms also mean 'nose.' Now, a lexical connection between 'nose' and 'prow of canoe' is also reported to be common in Austronesian languages (Blust 2009: 314), as suggested by the presence in Hawaiian in the present sample, and occurs also in Australia (Schebeck 1978: 175). The reason why this pattern is not detected in the sample is likely that the respective words in Austronesian languages do not simultaneously also cover 'beak' ('nose' itself is not among the meanings presently surveyed). Thus this case is not damaging to the validity of the sample, but it shows that still some caution is required before the findings are interpreted as being incontrovertible facts. In spite of this, a criterion that is in danger of being too lax rather than too stringent is opted for because rather than sorting out too many potentially areal cases and to report only a smallish list, it seems more desirable at this stage of research to provide a more comprehensive overview of preliminary candidates for areal lexico-semantic patterns in the lexicon as a resource for further research, even if some of them might turn out to be spurious. In spite of the shortcoming of not being able to offer statistical backup of the patterns, this method still has one big advantage over previous studies on areality in the lexicon, namely that it allows assessing areal patterns against solid cross-linguistic evidence from the entire world. In addition, it satisfies, in spite of its departure both from traditional accounts and from Bickel and Nichols's statistics-based approach, Masica's (2001: 207) programmatic statement as to the goals of areal linguistics: "[p]rimarily, areal linguistics should mean the study of significantly non-random distribution of linguistic features in space – first of all the facts and if possible the reasons behind them" (emphases removed).

In addition, the method used also by definition excludes the possibility that one and the same lexico-semantic pattern is diagnosed as being characteristic of more than one area at the same time. For the moment, the discussion is concerned with positive evidence, in spite of Masica's (2001: 215-216) valid reminder that linguistic areas can and should also be defined negatively, i.e. by absence of a certain feature (see § 6.4.3.14. for some casual notes on Eurasia in this respect).

The following tables show the lexico-semantic patterns that are diagnosed as being characteristic for the areas defined above, starting with Africa. Throughout, the first column states the lexico-semantic association, the second identifies the number of languages in the area with the pattern before the dash, and the number of languages outside the area (if any) after the dash. In the third column, the languages within the area participating in the pattern are named, and after them, the language(s) featuring the association

outside the area is/are stated in parentheses. The fourth column identifies whether the pattern is one of recurrent constituents in morphologically complex expressions (C), whether it is due to colexification (CL), or whether the pattern is mixed (M), i.e. the same lexico-semantic association is realized in at least one language as a morphologically complex expression and in at least one other by colexification. Such a situation should not be surprising, given the close ties between morphologically complex terms and colexification that have been emphasized throughout this study. The rightmost column is an important one: here references to the literature are stated. As already alluded to above, it is by no means the case areal patterns in the lexicon have entirely escaped attention of scholars so far. However, they are often mentioned en passant in other contexts (for instance, for the New Guinea area typically in discussions of difficulties in the applicability of lexicostatistics), and this column is intended to bundle previous statements in the literature that correspond to the detected patterns on the basis of the sample, and, importantly, to provide, where applicable, additional information on the historical emergence of the pattern and the cultural underpinnings that are likely responsible for them. Furthermore, the column also identifies statements in the literature where the lexico-semantic pattern is mentioned outside of the area for which it is presently detected, that is, where such information is available, it serves to flag patterns that may be spurious. Since the present method of assessing areality cannot be anything more than preliminary, patterns that may be areal, but on a smaller scale, are mentioned in prose casually after the tables, in particular if suggestions to this effect can be found in the literature, but relevant patterns fail to be diagnostic for the larger scale areas under investigation. As a final note, the analysis does not take into account the cross-linguistic differences in motivated terms between languages: it is conceivable that a language in which they are frequent has a greater chance of participating in areal patterns simply by virtue of having more motivated terms than another.

6.4.3.2. Areal Patterns of Africa

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'seed' – 'kind'	3-0	Efik, Hausa, Koyraboro Senni	CL	
2.	'sunrise' – 'sprout'	3-1	Hausa, Khoekhoe, Noni - Khalkha	C	

table 18: areal lexico-semantic patterns of Africa

There are a number of other associations only found in the present sample in two unrelated languages of Africa. For instance in Hausa and Khoekhoe, there is a metaphorical transfer from 'branch' to 'descendants,' and in Hausa, as well as the Niger-Congo languages Mbum and Yoruba, there are complex terms for the 'eclipse' featuring constituents

meaning 'to catch.' In Khoekhoe, the 'Adam's Apple' is called *dom-ǃkhom-s* 'throat-bundle-3SG,' and similarly, in Yoruba it is *kókó-ọ̀fun* 'lump-throat.' Hausa and Yoruba colexify 'buttocks' with 'root,' and Hausa and Buli colexify 'guts' with 'inner tube of tire.' In Hausa and Katcha, perhaps by functional contiguity, words for 'ball' also denote the 'nut of the dum-palm.' There is an association between 'cheek' and 'bag' or 'sack' in Mbum and Ngambay, and in Hausa and Swahili 'dawn' is colexified with 'morning prayer.'

6.4.3.3. Areal patterns of Europe

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'bay' – 'breast'	2-0	Basque, Greek	CL	also present e.g. in German, where it is a loan translation of Lat. *sinus maritīmus* and attested from the 17th century onward (Kluge 2002)
2.	'estuary' – 'sea'	2-0	Basque, Welsh	C	
3.	'house' – 'company, firm'	2-0	Basque, Greek	CL	

table 19: areal lexico-semantic patterns of Europe

Another putative European association is that between 'Milky Way' and 'milk,' attested in the present sample in Basque, and by a semianalyzable term in Greek. At least in German, the term is a 17th century calque from Latin *via lactea* (Kluge 2002); Latin may itself have been influenced by Ancient Greek. Likewise, a pattern having probably originated in Europe (though it may have Semitic origins) is the association between 'Adam's Apple' and 'apple,' which is attested in the present sample in Basque, Greek, and Welsh, but also in Itzaj, where it was apparently calqued from Spanish.

6.4.3.4. Areal patterns of South and Southeast Asia

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'waterfall' – 'cliff'	3-1	Bwe Karen, White Hmong, Yay – Copainalá Zoque	M	

table 20: areal lexico-semantic patterns of Mainland Southeast Asia

Surprisingly given the amount of literature mentioning Southeast Asia as a host for areality, also in the lexicon, there is only one pattern emerging as a candidate for an areal association by employing the present methodology. This has something to do with the fact that presently not only Mainland Southeast Asia is considered, and it may be that the Southeast Asian mainland, if considered on its own and/or with inclusion of Japanese data, would yield a considerably higher number of areal associations. For instance, there are

terms for 'animal' containing elements meaning to 'move' and 'thing' in Mandarin, Vietnamese, and Japanese, probably due to widespread borrowing from Chinese. Moreover, in Manange and White Hmong, but also in Haida, there is a term for 'lightning' involving an element meaning 'to blink eyes' (Bauer 1992 argues "that the three meanings 'wink', 'lightning', and 'wave' comprise a Southeast Asian word family," but also mentions that this association might extend to Austronesian languages). Also with regard to 'lightning,' there are associations with spirits in White Hmong and Yay, and there are terms for this meaning involving a constituent meaning 'sword' in Sedang and Yay. Also, a characteristic of this area are terms for the 'eclipse' involving a constituent meaning 'to eat' and terms for animals; however, 'eat' as a constituent of relevant terms is also attested in other languages outside Southeast Asia (compare Appendix E, 35 and Matisoff 2004).

6.4.3.5. Areal patterns of Eurasia

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'flower' – 'picture'	2-0	Khalkha, Kolyma Yukaghir	CL	
2.	'forest' – 'taiga'	2-0	Khalkha, Kolyma Yukaghir	CL	
3.	'lightning' –'arrow'	2-0	Khalkha, Kildin Saami	C	
4.	'rainbow' – 'thunder'	3-0	Ket, Nivkh, Kildin Saami	C	See Räsänen (1947)
5.	'river bed' – 'deep'	2-0	Ket, Kolyma Yukaghir	C	
6.	'guts' –'end'	2-0	Abzakh Adyghe, Kolyma Yukaghir	C	

table 21: areal lexico-semantic patterns of Eurasia

6.4.3.6. Areal patterns of Australia

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'moon' –'snail'	2-0	Burarra, Gurindji	CL	
2.	'nest' – 'raft'	2-0	Nunggubuyu, Tasmanian	CL	
3.	'boat' – 'collamon'	2-0	Gurindji, Yir Yoront	CL	
4.	'pen' – 'poke'	2-0	Gurindji, Yir Yoront	C	
5.	'breast/milk' – 'Burton's legless lizard'	2-0	Burarra, Nunggubuyu	CL	
6.	'night' – 'sleep'	2-0	Ngaanyatjarra, Nunggubuyu	CL	a fine example of actual-potential polysemy

table 22: areal lexico-semantic patterns of Australia

6.4.3.7. Areal patterns of New Guinea

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'animal' – 'pig dog'	3-0	Kyaka, Sentani, Takia	C	See Wälchli (2005), Farr (2001: 124) on Korafe.
2.	'egg' – 'nut'	4-0	Dadibi, Kaluli, Kyaka, Takia	CL	See Brown (1977: 299) on Eleman.
3.	'table' – 'floor'	3-1	Baruya, Dadibi, Toaripi, (Hawaiian)	M	
4.	'bone' – 'strong/strength'	3-1	Baruya, Kwoma, Waris, (Ngambay)	C	See Aikhenvald (2007: 549) on Manambu, which has been in contact with Kwoma

table 23: areal lexico-semantic patterns of New Guinea

Another New Guinea pattern mentioned in the literature for a particular language (Farr 2001: 126 on Arop-Lokep) is that of having terms for 'cloud' that can be literally translated as 'faeces of the wind.' Alongside Arop-Lokep, this pattern is found in another Austronesian language, Takia, that is spoken in close proximity. However, it is also found in Toaripi, spoken near the Southern Coast of New Guinea, and thus quite far removed from the other languages where it is known to occur. According to Brown (1977: 299), here the complex term replaces an older inherited Eastern Eleman word. Given the high idiosyncrasy of the pattern, it seems quite unlikely that the cases are independent of each other, and thus one would expect the association to be found in other New Guinea languages as well. Colexification of 'seed' and 'egg' is, as noted by Laycock (1970: 1141), indeed very common in the New Guinea area, but it also occurs in some other languages. Furthermore, there is an apparently widespread cultural association between the 'womb' and 'netbags' in New Guinea (see for instance Stewart and Strathern 1997). This is confirmed by the sample on the lexical level for two New Guinea languages, but the association is also present in Buin and Burarra.

6.4.3.8. Areal patterns of Oceania

There is just one association common enough to be diagnosed as areal according to the present methodology, namely that between 'star' and 'starfish,' which occurs by colexification in Buin, Lavukaleve, Rotokas, and Bislama.

6.4.3.9. Areal Patterns of Eastern North America

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'animal' – 'move on earth'	2-0	Blackfoot, Lakhota	C	
2.	'straw' – 'hat'	2-0	Pawnee, Cheyenne	M	
3.	'house' –'teepee'	2-0	Comanche, Kiowa	CL	
4.	'train' – 'run'	2-0	Biloxi, Chickasaw	C	
5.	'eyeball' – 'round object'	3-0	Blackfoot, Chickasaw, Oneida	C	
6.	'Saturday' – 'younger sibling'	2-0	Biloxi, Chickasaw	C	See Brown (1999: 26)

table 24: areal lexico-semantic patterns of Eastern North America

There are terms for 'Saturday' involving constituents meaning 'Sunday' and 'small, little' in Cheyenne and Kiowa, but taken by themselves, both 'Sunday' and 'small, little' recur as constituents in terms for 'Saturday' in other configurations (see Appendix E, 158).

6.4.3.10. Areal Patterns of Western North America

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'beak' – 'peck'	3-0	Upper Chehalis, Ineseño Chumash, Nuuchahnulth	C	perhaps present in Hani as well
2.	'branch' – 'knot in tree/knot in wood'	4-1	Carrier, Upper Chehalis, Lake Miwok, Nez Perce, (Central Yup'ik)	CL	
3.	'dew' – 'wet/moist'	3-0	Upper Chehalis, Ineseño Chumash, Kiliwa	M	
4.	'thorn' – 'sticker'	3-1	Cahuilla, Lake Miwok, Wintu, (Pawnee)	C	
5.	'car' – 'move self'	3-1	Carrier, Nez Perce, Kashaya, (Lakhota)	C	see some remarks in Appendix E, 77

6.	'car' – 'machine'	4-1	Ineseño Chumash, Kashaya, Lake Miwok, Wappo, (Kildin Saami)	CL	all terms in Western North America borrowed from Spanish

table 25: areal lexico-semantic patterns of Western North America

6.4.3.11. Areal Patterns of Mesoamerica

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'resin' – 'birdlime'	2-0	Itzaj, San Lucas Quiaviní Zapotec	CL	
2.	'rope' –'lasso'	4-1	Itzaj, San Mateo del Mar Huave, Xicotepec de Juárez Totonac, Copainalá Zoque, (Cubeo)	CL	

table 26: areal lexico-semantic patterns of Mesoamerica

The paucity of the results is surprising, since there is, as noted already above, a series of publications that demonstrate that Mesoamerica forms a linguistic area, also when it comes to the lexicon. The present sample confirms the association between 'ashes' and 'lime' noted by Smith-Stark (1994), which occurs here in Itzaj and Copainalá Zoque, but also in Tetun by the analyzable term *ahu-metan* 'lime-black.' As already noted earlier, some of the features mentioned in Campbell et al. (1996) and Smith-Stark (1994) do not emerge as being characteristic of Mesoamerica under the present method of assessing areality. The fact that a region of the world which has been studied in great detail from the point of view of areal phenomena in the lexicon yields a very small number of results when compared with the other areas suggests that there is a lot more interesting work to be done for other areas of the world in areal linguistics.

6.4.3.12. Areal Patterns of South America

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'beak' – 'prow of canoe'	5 (6)-1	Aguaruna, Bora, Guaraní, Huambisa, Piro, Wayampi, (Hawaiian)	M	Also common in Austronesian according to Blust (2009: 314), as suggested by the presence in Hawaiian
2.	'cloud' – 'smoke'	7 (8)-3	Arabela, Bora, Cashinahua, Cavineña, Cayapa, Hupda, Maxakalí, Tsafiki, (Buin, Sedang, Nez Perce)	M	
3.	'dust' – 'smoke'	5-1	Carib, Guaraní, Hupda, An-	M	

			cash Quechua, Tsafiki, (Fijian)	
4.	'knife' – 'machete'	7-3	Bora, Cavineña, Chayahuita, Cubeo, Hupda, Jarawara, Yanomámi – Meyah, Sko, Basque	M
5.	'neck' – 'nape'	9 (10)-1	Abipón, Aguaruna, Arabela, Aymara, Cubeo, Piro, Ancash Quechua, Tehuelche, Toba, Wayampi, (Hani)	M

table 27: areal lexico-semantic patterns of South America

There are at least two further conspicuous patterns that occur on a smaller scale in South America: one is to have a semantic extension of the word for 'beard' to also mean 'antennae of an insect.' This is found in the sample in three languages spoken on the eastern slopes of the Andes: Arabela, Cavineña, and Piro (as well as in Mesoamerica in Xicotepec de Juárez Totonac). Furthermore, South America is a hotspot for complex terms for 'mouth,' either as 'teeth hole,' 'speak hole' or 'language hole.' Such terms occur in Jarawara, Maxakalí and Tsafiki, in other words, without clear confinement on a smaller scale within South America. There are also a number of languages in South America with complex terms for 'dew,' either on the basis of 'urine' or 'saliva' (for the latter in the Vaupés area compare Epps 2007: 285, see also Zamponi 2009: 590 for data from Maipure), but ones on the basis of 'urine' also occur in Australia. Also common is a lexical association between 'stomach' and 'faeces,' either by polysemy, as for instance in Bororo, or by way of morphological complexity, as in Piro, where *hitš̲ka-mapa* contains *hitš̲ka* 'faeces' and *mapa* 'bag, bladder.'

6.4.3.13. Larger linguistic areas?

6.4.3.13.1. Introduction. As stated in the general introduction to the discussion of areality in semantic associations, on the basis of the present sample it is only possible to deal with quite large areas due to the insufficient coverage for small geographically restricted regions of the world (for instance, there is only one language, Greek, which participates in the Balkan Sprachbund). However, there is some evidence from evaluation of the areal clustering of semantic associations that even larger linguistic areas exist than those discussed so far, namely in those cases when there is a notable areal bias to the distribution of a given feature, which however fails to be diagnosed as potentially areal for any of the areas looked at so far, not because the pattern occurs with some frequency at random scattered elsewhere in the world, but because its representation is relatively strong also in areas geographically adjacent to that in question. This is the case, first, for Eurasia as a whole, including Europe, second, for Mainland Southeast Asia and Oceania broadly (including New Guinea), third, for the Old World, that is, Eurasia and Africa, and finally, for North America as a whole. The following discussion presents the relevant evidence that emerges when the methodology introduced in § 6.4.3.1. is applied to these macro-areas.

6.4.3.13.2. Putative semantic associations characteristic of Eurasia as a whole

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'fog' – 'darkness'	2-0	Welsh, Khalkha	CL	
2.	'Saturday' – 'Saturn'	3-1	Badaga, Khalkha, Welsh, (Japanese)	M	
3.	'Saturday' – 'unique'	2-0	Basque, Abzakh Adyghe	C	

table 28: putative areal patterns of Eurasia as a whole

6.4.3.13.3. Putative semantic associations characteristic of Mainland Southeast Asia and Oceania.

	Lexico-Semantic Associations	Number of Languages	Languages	Type	Comments
1.	'foam' – 'saliva'	9 (10) - 3	Baruya, Buin, Kwoma, Lavukaleve, Muna, Ngaanyatjarra, Sko, Tasmanian, Bislama, Lenakel, (Badaga, Lesser Antillean Creole French, Tsafiki)	M	
2.	'beard' – 'feather'	7 (8)-3	Berik, Kwoma, Meyah, Toaripi, Bwe Karen, Hawaiian, Lenakel, (Efik, Chickasaw, Toba)	M	

table 29: putative areal patterns of Mainland Southeast Asia and Oceania

A common pattern in both Southeast Asia and Oceania is to have complex terms for 'sun' on the basis of 'eye' (Urban 2010b). Furthermore, many languages of this area, namely Anggor, Muna, Sahu, Sedang, and Bislama colexify 'bark' with 'husk' (this association is also heard of in Niger-Congo languages), and Kwoma, Toaripi, Hawaiian, and Samoan have a lexical association between 'kidney' and 'fruit,' while there is a semianalyzable term in Bwe Karen. Also common in the broader area of New Guinea and Oceania, but rare though not unheard of elsewhere, are associations between 'egg' and 'fruit' and 'egg' and 'seed' (for New Guinea, compare Laycock 1970: 1141, and Brown 1977: 299 for Eleman more specifically). Holmer (1966) draws attention to associations between 'place' and 'time' in Oceania more generally; two languages in the sample behave accordingly, Kyaka and Yir Yoront, but this is not enough to satisfy criterion (ii).

6.4.3.13.4. Putative semantic associations characteristic of the Old World.

	Lexico-Semantic Associations	Number of Languages	Languages	Type
1.	'house' – 'lineage'	5-1	Buli, Rendille, Abzakh Adyghe, Badaga, Basque, (Wintu)	CL
2.	'pen' – 'feather'	5-1	Efik, Khoekhoe, Basque, Nivkh, Kildin Saami, (Hawaiian)	CL

table 30: putative areal patterns of the Old World

Also common, though not common enough to satisfy criterion (ii) of the present methodology are the following associations: colexification of 'animal' with 'livestock,' 'sky' with 'god,' 'star' with 'blaze,' 'tail' with 'buttocks,' 'ball' with 'bullet,' 'clock' with 'bell,' 'street' with 'journey,' 'heart' with 'boldness, courage,' and that of 'neck' with 'neck of vessel;' see relevant sections in Appendix E.

6.4.3.13.5. Putative semantic associations characteristic of North America

	Lexico-Semantic Associations	Number of Languages	Languages	Type
1.	'clearing' – 'field/meadow/lawn'	5-2	Carrier, Kashaya, Lake Miwok, Pawnee, Tuscarora, (Central Yup'ik, Badaga)	CL
2.	'mirror' – 'window'	6-1	Upper Chehalis, Kashaya, Kiowa, Pawnee, Tuscarora, Wintu, (Fijian(	CL
3.	'bed' – 'lie/lie down'	9 – 4 (8)	Carrier, Upper Chehalis, Cheyenne, Ineseño Chumash, Haida, Kiliwa, Lakhota, Wintu, Yuki, (Efik, Kwoma, Jarawara,Fijian, Hawaiian, Malagasy, Rotuman, Tetun)	M

table 31: putative areal patterns of the North America

In some languages of North America, namely Carrier, Cheyenne, and Lakhota, there is an association that likens the 'rainbow' to a 'snare' (see Hall 1997: 56). Notable are also the associations between thunder and a mythological thunderbird (compare Eells 1889: 335). This pattern is apparently most widespread in the Northwest, occurring in Carrier and Upper Chehalis (see also Sapir 1916/1949), but clearly extends further west and south (e.g. Chamberlain 1890 on Algonquian). Interestingly, the association is also found in Miskito, and, outside of the Americas, in Waris, spoken in New Guinea. Other perhaps notable patterns in North America are complex terms for 'sky' involving a constituent meaning 'blue' or colexification of these meanings, which is the case in Biloxi, Upper Chehalis, Cheyenne, Oneida, Tuscarora, but also in Hawaiian. Also, in three languages of North America, Lakhota, Nuuchahnulth, and Pawnee, there is an association between 'paper' and 'cloth.'

6.4.3.13.6. On the possibilities of the emergence of macro-areality in semantic associations. How could these patterns, if they are real and can be substantiated, have arisen? Is there any plausible scenario emerging from the history of the relevant regions that could account for these distributions, however abstract it may be?

There is at least one of the associations identified as being characteristic of a large continent-sized area, Eurasia, for which the history is documented: that of 'Saturday' with 'Saturn,' and it is illuminating to discuss the emergence of this association as a case study of how large-scale areality in the lexicon can emerge. Eurasia is the host of a number of commonalities in the lexicon, and it is here that one can probably best trace at least some of its developments. The continent is a spread zone in terms of Nichols (1992), which is to a great extend due to its geography. Eurasia does not feature major geographical boundaries to west-east travelling and the geography thus provides a fostering environment for cultural and linguistic contact. One instance of these is the long-established west-east trade along the Silk Road (Beckwith 2009). Nichols and Peterson (2005) and Bickel and Nichols (2009) have pointed to linguistic outcomes of this situation. Consonantism patterns in pronouns involving a bilabial nasal in the first person pronoun root and an alveolar stop or affricate in the second person pronoun root, for instance, are hypothesized to be the result of precisely the aforementioned cultural continuities that link the whole continent since ancient times. At a later point of time, after the initial establishment of the east-west trading networks the origins of which go back to the Bronze Age, the conquest of Persia by Alexander the Great and his advance further east brought about a fusion of the Ancient Greek culture with that of the Middle East and Asia, renewing pan-Eurasian cultural ties. This is known as the Hellenistic period, commencing in the third century BC, and it is here that at least one of the lexico-semantic patterns characteristic of Eurasia has its origin, namely that of the association between 'Saturday' and the planet 'Saturn.' The practices of naming the days of the week after the seven planets (including sun and moon) visible with the bare eye is of Mesopotamian origin, and towards the end of the Hellenistic period, the system was adopted in the Hellenistic world, and names were given to the days of the week on the basis of the Greek names of the planets (the precise line of development is a little more complicated, see Cumont 1935 for details). Subsequently, the system established itself in Greek-controlled Egypt and in the Roman World by the first century BC (Sarton 1959), from where it in turn made its way to the vernaculars of Europe due to Roman dominance. It spread to the Indian subcontinent along with Hellenistic astronomy, and coexisted there with the indigenous Indian calendar based on lunar cycles (Markel 1995). Its presence in Japanese is of relatively recent origin, it being a loan translation from Chinese in the context of Sinicization that is first attested in Japanese in 1444; Mandarin as spoken today does not use the planetary-based model term anymore (Schmidt 2009).

Some space was devoted to this particular case in spite of the danger to drift into a *Wörter und Sachen*-style hybridization of linguistic and encyclopaedic cultural facts, because it allows one to catch a glimpse at the development of and the historical facts responsible for a particular feature common to the lexicon of languages spoken in a very large, continent-sized area. No similar historical account is available for the emergence of

other patterns, such as the apparently widespread Eurasian association between 'rainbow' and 'thunder,' and its origins therefore remain obscure, but their existence nevertheless demonstrate the long history of cultural and linguistic continuities over Eurasia that result in some similarities in linguistic structures over the entire continent.

As for potential still larger areas such as the Old World, discussion, unfortunately, must remain more speculative. Notably, though, Africa, in particular Northern Africa, has had historical ties with Eurasia throughout the historical period: first by way of the Ancient Egyptian empire, later, by Hellenistic influence in North Africa; still later, the spread of Arabic culture in medieval times to both Africa and Europe as well as colonization of the African continent by European powers may all have played a role in the shaping of individual commonalities in lexico-semantic associations. As for Southeast Asia and Oceania, a candidate for bringing about areality that immediately comes to mind is the Austronesian spread from Southeast Asia to Oceania, which is known to have resulted in prolonged language contact with notable effects on linguistic structure in New Guinea in particular. All this, including the putative large-scale associations in particular, requires empirical substantiation, but at least, from the point of view of history, the possibility of the emergence of such large-scale patterns does not seem to be ruled out entirely.

6.4.3.14. A short note on negative evidence, concerning Eurasia

While in general the investigation is concerned with positive rather than negative evidence for assessing areality, there is nevertheless evidence that Eurasia also forms a large linguistic area that is constituted jointly by the striking absence or rarity of some lexico-semantic patterns that are so common otherwise that their presence can be considered the norm rather than the exception. These include:

(i) rarity of the association between 'milk' and 'breast' exceptions: Ket *mamul* 'breast-water' and Kolyma Yukaghir *ibiši* 'milk, breast, nipple'
(ii) absence of complex terms for 'lip' of the type 'mouth-skin'

With respect to the association between 'skin' and 'bark' a similar west-east cline is observable: Eurasian languages which have the association tend to be spoken in the east rather than in the west. The westernmost representative in the sample is Abzakh Adyghe, followed by Ket, Kolyma Yukaghir and Japanese (in Chukchi, the current form *itqilyən* is lexicalized from **ut(tə)-qulyə(n)* 'tree-skin' and non-transparent synchronically).

6.4.3.15. 'Things'

This section takes up a topic alluded to in various places in the discussion so far. The phenomenon at stake is that instead of derived terms, in some languages there is a high number of analyzable terms of the lexical type, in particular in the domain of artifacts, involving a constituent simply meaning 'thing.' Prominent among these is Cheyenne, and two examples from this language are in (2.).

(2). a. *šéešestôtse* /šéeše-hestôtse/ 'lie-thing' = 'bed'
b. *he'enénestôtse* /he'e-nén-hestôtse/
'female-nurse-thing' = 'nipple'

Table 32 provides data on the occurrence of fully analyzable terms of the lexical and derived type involving a constituent glossed as 'thing' within the meanings under investigation, separately for the domain of artifacts, since this is where most such terms accumulate and terms in other domains. Figure 15 is a plot of the distribution of such terms in the entire (EXT-2) sample.

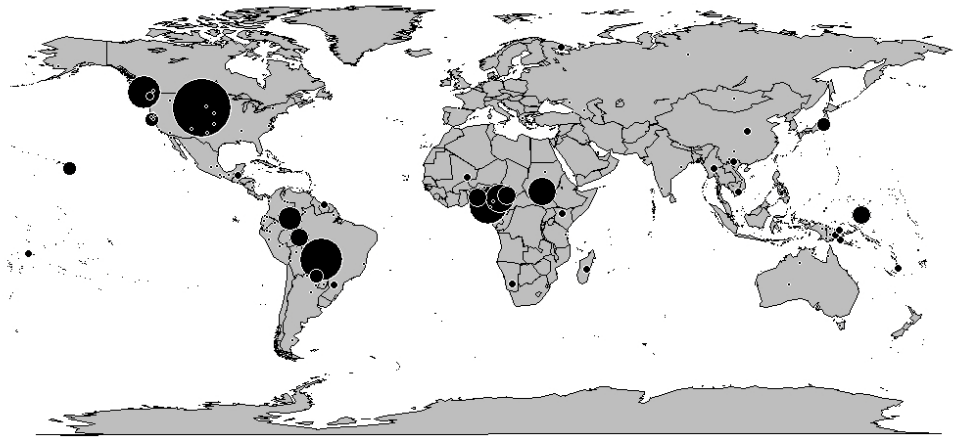


fig. 15: terms involving a constituent 'thing' in the full (EXT-2) sample

Language	'Thing'-terms in Artifacts	'Thing'-terms elsewhere
Efik	2	5
Katcha	2	3
Khoekhoe	0	1
Mbum	5	0
Ngambay	3	0
Rendille	0	1
Yoruba	1	2
Dadibi	1	0
Toaripi	1	0
Japanese	0	2
Kildin Saami	1	0
Cheyenne	9	2
Itzaj	0	1
Kashaya	2	0
Nuuchahnulth	2	4
Quileute	1	0
Bororo	4	4

Guaraní	1	0
Hupda	4	0
Jarawara	0	3
Kaingang	1	0
Lengua	2	0
Hawaiian	1	1
Kapingamarangi	1	2
Bwe Karen	0	1
Lenakel	0	1
Malagasy	0	1
Mandarin	0	1
Samoan	1	0
Takia	1	0
Vietnamese	0	1
Yay	0	1

table 32: terms involving a constituent meaning 'thing' in the sampled languages

Two words of caution are in order: the first pertains to the question as to how accurately the sampled meanings represent the situation in the entire lexicon. For instance, although not figuring prominently in the selection of vocabulary items on the wordlist and being sometimes redundant from a purely semantic point of view, compounds with *mar* 'thing' are also very frequent in Meyah, for instance *márféb* /már eféb/ 'thing string' = 'string,' *márfók* /már ofók/ 'thing bud' = 'flower,' *már éij* 'thing throw' = 'trash' (Gravelle 2004: 53; see also § 4.5.1.4.1). Second, table 32 does not take into account the number of available terms per language for the investigated meanings, but merely counts terms with the structure of interest.

Still, bearing in mind these points, the distribution does not appear to be entirely fortitious: notable is an areal bias in Africa and the Americas that is significant at p = .01203 when performing a Kruskal test using the standard six-way areal breakdown of the world (χ^2 = 14.6382, df = 5, only using languages of the statistics sample to control for inheritance).

However, there is more to be said: 'thing'-terms, both for artifacts and non-artifacts, are found with a high frequency in languages with a rather modest apparatus of nominal morphology. This is true of most African languages in the sample, but it is equally true of Macro-Gê languages in South America, Vietnamese in Southeast Asia, and of Oceanic, in particular Polynesian, languages of Oceania. One might be inclined to conjecture that in such languages forming complex terms on the basis of 'thing'-words replaces a missing nominalizing morpheme (and note Moser's 2004: 133 discussion of *né* 'thing' in Ngambay as a deverbal nominalizer as well as the dictionary gloss for Nuuchahnulth *-mis* "thing, used as a nominalizer" and the same situation which obtains in Bwe Karen; terms involving these elements were taken into account in table 32), in other words, that 'thing'-terms are circumlocutory transcategorial operations in languages that lack morphological means to do so (indeed, a grammaticalization cline 'thing' > 'nominalizer' is perfectly con-

ceivable, although not mentioned by Heine and Kuteva 2002). Notably, though, languages in which 'thing'-compounds accumulate do not necessarily lack noun-deriving morphology, in particular a general nominalizer, altogether, so this cannot be the whole story. Some languages with many 'thing'-terms typically lack dedicated derivational morphemes for instrument and locative nominalization (or employ these so rarely that they do not figure in the database), but this generalization is far from going the whole way as well: for instance, Kapingamarangi and Malagasy feature prefixes *k-* and *faN-* respectively for forming instrument nouns, and there are other languages for which this explanation is not available either.

However, there is some evidence, although limited, that one is dealing, at least in non-Andean South America, with an areal phenomenon. In this regard, Jarawara *jama* 'thing' deserves special mention. It can and apparently very frequently does refer to any object when the context allows identifying the referent unambiguously. Dixon (2004: 540) states that

> *[j]ama* is the generic term par excellence. The normal gloss offered is 'thing.' It can be used for any new object or foodstuff, for which a name is not known, or simply as a vague term instead of employing a more specific name. *Jama* can be used for 'season' or 'time' ... *Jama* can be used to refer to the forest (a more specific name is *jama.kabani*; no etymology is known for *kabani*) or to game in the forest, or to the spirits of the forest. *Jama* can also be used to refer to fishes in the river. A PN [possessed noun] is generally used with a free noun; *jama* is often the 'dummy head' with a PN. For instance, free noun X plus PN *abe/ebene* is 'living being associated with X', e.g. the people inhabiting a place called X. ... Insects in general can be described as *jama abe*, literally 'creatures associated with a thing', using *jama* in its most general sense. ... The wide range of meaning and use of *jama* is exhibited in one sentence from a text which has *jama* as its A argument and also jama as its O argument (but with different reference): *jama jama firi kasa* 'the thing (here, lightning flash) fully illuminates the thing (here, a dead body)' ...

Alongside being capable of referring to the weather, especially salient seems to be the use of *jama* to refer to a specific time or place:

(3.) a. *jama hiwa-bote ama-ke*[16]
thing(f) be.hot-VERY be-DECf
'the weather (lit. thing) is very hot' (adapted from Dixon 2004: 337)

[16] Glosses: 1EXC '1st person non-singular exclusive (excluding addressee),' A 'transitive subject,' AUX 'auxiliary,' CINT 'content interrogative,' DEC 'declarative mood,' f 'feminine', IMMED 'immediate mood,' NOM nominalization,' PERI peripheral,' S 'intransitive subject,' sg 'singular.'

b. *himata jama jaa ti-ka-ma-ri-be?*
what thing(f) PERI 2sgS-in.motion-BACK-CINTf-IMMEDf
'when (lit: at what thing) will you return?'
(adapted from Dixon 2004: 409)

c. *jama jabo-ke faha otaa kii ni kaaro*
thing(f) be.far-DECf waterf 1excA search.in AUX+NOM PERIf
'the place where we fish is far off (lit. the thing (place) is far off, we search the water at it' (adapted from Dixon 2004: 500)

Jama is also very frequently employed for word-formation tasks, and the lexemes formed with it appear to have varying degree of conventionalization. Instrument nouns are frequently formed by combining *jama* with a reduplicated form of a verb to denote items of acculturation, such as *jama ho-howe* 'thing RED-clean.out' = 'rake' (adapted from Dixon 2004: 534); none of them is found in the database, though. In this function, *jama*, given its extremely vague semantic content, approaches a nominalizer in function. *Jama* is equally readily combined with another noun for the purpose of word formation, as seen e.g. in *jama soki* 'thing be.dark' = 'night' and, conversely, *jama wehe* 'thing light' = 'day.'

Interestingly, in Bororo, a similar, although not identical situation is encountered. Bororo *boe* is prominently used as the autonym with which the Bororo refer to themselves:

(4.) *Boe e-tu-re*
Bororo 3PL-go-NEUTRAL
'The Bororos left.' (Crowell 1979: 227)

However, as Crowell (1979: 226) remarks, the term, like Jarawara *jama*, "occurs with great frequency, along a scale of specificity." *Boe* can also be used to refer to other Indians or people, or it can (or must) be translated by 'thing.' Parallel to the range of use of *jama* in Jarawara, *boe* is also used when talking about the weather or about time:

(5.) a. *Boe uru-re*
thing hot-NEUTRAL
'It's hot'

b. *Boe xo-re*
thing black-NEUTRAL
'It's dark (or night).' (Crowell 1979: 226, glosses adapted)

While there may be an areal factor in play (note also the parallelism with respect to the structure of the expressions for 'night' in Bororo and Jarawara), there also is an alternative explanation: high frequency of a semantically underspecified noun, presupposing a high amount of implicitly shared cultural and real-world knowledge, may be a symptom of languages used primarily for intra-group communication (Thurston 1989, Wray and Grace

2005). As Dixon (2010: 301) notes, "[i]n contrast, there are languages which lack anything resembling a generic noun 'thing'." In fact, the behavior of languages with respect to this seems to be not unconstrained, but might rather be accountable for by sociolinguistic factors.

6.4.3.16. General Discussion.

A clear division between facts and historical explanation is important for the context of the present study. The question that arises now after surveying the patterns is whether all of them necessarily must have a historical explanation. In spite of the fact that with a broad geography-based approach to areal linguistics, one should not necessarily expect isoglosses as in geographically more restricted areas or even dialect geography (Bickel and Nichols in press), some of the patterns found in continent or subcontinent-sized areas, for instance North America, have a very discontinuous geographical distribution while the overall frequency is not necessarily very high (though, by definition, higher than 15%). While it is not impossible that they indeed are indicative of historically grown areality that extends across the continent, this need not necessarily be so. This is partly due to the sparseness of historical data for some areas of the world that would allow coming up with more specific contact scenarios and due to the lack of research of areality in the lexicon that is concerned specifically with large areas. However, this statement should not distract from the fact that some of the diagnosed patterns may simply be spurious. In connection with this, it is notable that there are also a number of terms for items of acculturation, in particular in the Americas, that are diagnosed as being areal. Rather than being an indication for continent-wide diffusion, these are by far more likely to be relatively uniform responses to items of acculturation that were previously unknown (however, some diffusion of semantic associations is historical times seems to have happened. For instance, Siouan and Algonquian languages feature terms for 'distilled spirits' literally translatable as 'fire-water' according to Rankin 2003: 193, a pattern they share with Dene Sųłiné as reported by Rice to appear. Similarly, languages of the Southeastern United States feature terms translatable as 'bitter-water' for 'whiskey' according to Brown 1999: 146, table 11.1.). Obviously, then, these are of little to no value for an areal linguistics that seeks to identify diffused structural traits in neighboring languages due to language contact. However, they are still amenable to a somewhat less interesting, because obvious, historical explanation in that the artifacts they designate, such as 'pens' or 'tables,' were previously unknown in the areas where the morphologically complex terms for them occur.

A general observation is that there is some variation between areas whether the found patterns are more frequently due to colexification or due to parallelism in morphological structure. While, for instance, the patterns found in the Old World are characteristically due to colexification (this is not an absolute statement but a generalization only: the areal patterns found in Europe, to the contrary, are entirely due to calquing), those in the Americas tend to be found in parallelisms in morphologically complex lexical items – a finding that is in line with the general distribution of morphological complexity on a macro-areal scale: comparably low degree of analyzability in the Old World, comparably high degree in the New World. All in all, there are 21 patterns of colexification and 17

patterns in morphologically complex terms that are diagnosed as being associated with one of the tested areas under the present method. In other words, if the diagnosed patterns are genuine, the results show that areal influence and convergence in the semantics of lexical items alone, without replication of morphological structure, is roughly equally frequent as calquing in the traditional sense, and thus seems to be of value for research in linguistic areality and language contact generally.

6.5. CROSS-LINGUISTICALLY COMMON LEXICO-SEMANTIC ASSOCIATIONS

One value of the discussion of individual concepts in Appendix E noted by Blank (2003) is that such a list of cross-linguistic associations allows one to predict to some extent that, should a neologism be coined in some language of the world, speakers are likely to choose one of the conceptualization strategies listed there. However, another value in the opposite direction that has been noted early on in the literature as discussed in chapter 2 is that, should terms be etymologizable, they are also likely to exhibit one of the lexico-semantic associations found synchronically in other languages, and in this sense, the issue is intricately linked with questions of diachrony.

Having discussed structures that are probably areal in their distribution, one is left with two types of distributions in lexico-semantic associations: those that are very rare cross-linguistically, occurring for instance in only two or three languages in the sample without any appreciable hotspot anywhere, and those that are so common that they are encountered in very many parts of the world in very many languages. The purpose of this section is to present the latter. The data for the entire (EXT-2) sample are used for this purpose, on the one hand to allow for a maximally inclusive dataset for this explorative investigation, on the other since, as discussed in § 6.4.2., there is little (though not no) evidence for inheritance and genealogical stability of most patterns anyway, a fact which justifies this decision.

However, before it is possible to do that, one needs to decide how to sort out the areally unrestricted rare patterns from the areally unrestricted common patterns. This may seem like an odd question to discuss explicitly, but the two types of data are not separated from one another by some sort of intrinsic criterion, but form a continuum. As with the very first typology in § 4.2. pertaining to the relative prevalence of analyzable terms of the lexical and derived type, the relative frequency of each association (both by morphological analyzability and colexification) is calculated, and then divided by the representation score for the meaning it pertains to in order to account for the variable recoverability of relevant terms from the consulted sources. All in all, there are 1,892 lexico-semantic associations (few of them are doublets since both associated meanings are on the 160-items list used). Figure 16 is a histogram showing the frequency distribution.

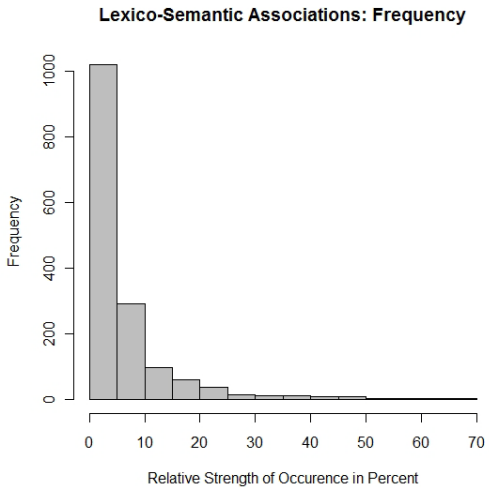


fig. 16: relative strength of occurrence in percent for recurrent lexico-semantic associations

The distribution is approximately negatively exponential: the vast majority of lexico-semantic associations is relatively rare, and there are only very few extremely frequent associations. The most common is that between 'eyeball' and 'eye,' found in 94.34 per cent of sampled languages for which data are available, mostly because languages have analyzable terms for 'eyeball' featuring a wide variety of mostly metaphor-based denominations for the concept which is contiguity-anchored by 'eye' (see Appendix E, 129 for details). One of the least frequent association is that between 'moon' and certain types of snails, found only in two languages of Australia, a pattern that may in fact be an areal association peculiar to Australia (there are very many associations present in only two languages; this one comes out as one of the least frequent since lexical equivalents for 'moon' are available for almost all sampled languages, and hence its relative strength is lower than if a pattern is attested twice, but for a meaning for which few terms could be extracted; note also that an occurrence in two languages is the lowest possible figure here, since all associations occurring only once are discarded according to the methodology described in § 3.7.3.2.).

Also parallel to the discussion of analyzable terms of the lexical and derived type in § 4.2., a division of the associations into four quartiles according to their frequency may be carried out, with everything in the fourth quartile defined (arbitrarily, but at least with a boundary that is generated out of the frequency counts themselves) as being a common association. This comprises all associations with a strength ranging from 7.14 per cent to the maximum (the abovementioned association between 'eyeball' and 'eye'). Of course, this means that there is still huge variation within this group, but there also is the advantage that for an exploratory investigation, the likelihood of missing a pattern that may in fact be more common than suggested by the present sample is greatly reduced. In absolute numbers, 485 associations are found within this range.

In the present sample, the most common associations are contiguity-based. Alongside the already mentioned association between 'eyeball' and 'eye,' it is also very common cross-linguistically to have analyzable terms for 'nostril' having a constituent 'nose' acting as the contiguity anchor (with little variation as to the meaning of the second constituent, which is most commonly 'hole, opening,' see Appendix E, 129). Complex terms for 'eyelid,' 'eyelash' and 'pupil' featuring, like those for 'eyeball,' a constituent meaning 'eye' are also very common. As for colexification, very commonly (each figuring among the 20 most common associations), 'moon' and 'month,' 'tree' and 'wood,' 'milk' and 'breast,' 'soil' and 'land, ground' generally, 'paper' and 'letter' or 'book,' and 'rope' and 'thread, string, twine,' or 'fibre' are colexified. Rather than listing each association here separately, Appendix D features a column in which common associations for each meaning as just defined (as well as putative areal patterns), should there be any, are listed as a resource for further research.

In fact, of the 476 common associations for which no areal hotspot was discernible according to the methodology used in § 6.4.3, roughly two thirds are mostly contiguity-based, and only about one third mostly metaphor-based, while for some of them several analyses are possible. From these figures alone, however, one cannot yet infer any reliable generalizations about the way languages conceptualize referents as revealed by motivated terms. This on the one hand due to the fact that many of the contiguity-based associations that are frequent are in fact only contiguity anchors for varying metaphorical transfers, each of which taken by itself is usually too infrequent to end up in the fourth quartile (terms for the 'eyeball' are an exception). On the other hand, even without taking into account this type of association, contiguity-based associations are more numerous among those that recur frequently in different languages (note that metonymy has come to be regarded by some Cognitive Linguists as the more important cognitive process when compared with metaphor, e.g. Taylor 2003: 126). Given this result, it is instructive to look in more detail at the frequent metaphor-based associations to see whether one can derive any generalizations as to the kinds of meanings for which they are common.

The following is a rough division for the sake of bringing order into the wealth of different associations (it is intended to only give a rough classification, and assignment of individual patterns may be debatable; also, not listed are associations for which several analyses are possible, for instance that between 'heart' and 'center, middle'):

(i) Analogies (biologically speaking, often homologies) in humans and plants or animals: 'bark' – 'skin' (59.42%), 'fingernail' – 'claw/talon' (48.94%), 'bark' – 'peel/rind/shell' (42.75%), 'feather' – 'hair' (35%), 'beak' – 'nose' (31.19%), 'breast' – 'udder/teat' (30.34%), 'feather' – 'fur/wool' (29.29%), 'bark' – 'hide/leather' (23.91%), 'skin' – 'rind/peel' (23.61%), 'breast' – 'nipple/teat' (23.45%), 'beak' – 'mouth' (21.1%), 'branch' – 'arm' (20.71%), 'nest' – 'house/home' (19.08%), 'skin' – 'shell' (18.75%), 'eyelid' – 'peel/rind/shell' (17.14%), 'eyelid' – 'bark' (15.24%), 'beak' – 'snout/muzzle' (14.68%), 'fingernail' – 'hoof' (14.18%), 'beard' – 'whiskers' (13.97%), 'eyelash' – 'fur' (13.18%), 'eyelash' – 'feather' (12.4%), 'eyebrow' – 'fur' (11.03%), 'nipple' – 'udder/teat'

(10.75%), 'branch' – 'hand' (10.71%), 'eyelid' – 'leather/hide' (10.48%), 'beak' – 'tooth' (9.17%), 'tendon' – 'root' (9.01%), 'branch' – 'leg' (7.86%), 'feather' – 'leaf' (7.86%), 'skin' – 'husk/chaff' (7.64%), 'bark' – 'scale' (7.25%)

(ii) Body parts and body liquids: 'tendon' – 'vein/artery' (49.55%), 'fingernail' – 'toenail' (27.66%), 'finger' – 'toe' (26.06%), 'phlegm' – 'saliva/spittle' (24.32%), 'vein' – 'nerve' (20.17%), 'snot' – 'phlegm/sputum' (15.57%), 'brain' – 'marrow' (13.53%), 'phlegm' – 'snot' (13.51%), 'nipple' – 'eye' (10.53%), 'uvula' – 'child/son' (10%), 'uvula' – 'vagina/clitoris' (10%), 'nipple' – 'head' (9.47), 'calf' – 'belly/stomach' (9.4%), 'ankle' – 'eye' (8.59%), 'eyebrow' – 'feather' (8.09%), 'semen' – 'pus' (7.58%), 'uvula' – 'tonsil' (7.5%)

(iii) Aerosols: 'steam' – 'smoke' (31.73%), 'fog' – 'cloud' (29.77%), 'steam' – 'fog/mist' (15.38%), 'fog' – 'smoke' (14.5%), 'cloud' – 'smoke' (7.86%)

(iv) Artifacts: 'ladder' – 'stairs/staircase' (26.09%), 'window' – 'door' (20.95%), 'glasses' – 'mirror' (18.6%) 'airplane' – 'boat/canoe' (15.63%)

(vi) Mythology-/culture-based associations: 'sky' – 'heaven' (30.77%), 'shadow' – 'soul/spirit/ghost' (27.74%), 'shadow' – 'reflection/mirror/image' (24.82%), 'heart' – 'feel/think' (18.06%), 'shadow' – 'image/picture/drawing' (20.44%), 'thunder' – 'god/spirit' (9.63%), 'heart' – 'soul/spirit' (8.33%)

(vii) Generalizing/narrowing: 'rope' – 'thread/string/cord/twine' (49.61%), 'mountain' – 'hill' (42.54%), 'coast' – 'edge/end/border' (35.63%), 'train' – 'wagon/vehicle' (35%), 'buttocks' – 'bottom/base' (23.39%), 'resin' – 'water/liquid/juice' (21.65%), 'mouth' – 'opening' (20.83%), 'ashes' – 'dust' (19.72%), 'dust' – 'dirt/rubbish/garbage' (19.12%), 'semen' – 'water/juice' (18.18%), 'chair' – 'furniture' (17.6%), 'valley' – 'gully/furrow/ditch/gorge/channel' (17.54%), 'lake' – 'lagoon' (17.36), 'skin' – 'surface/cover' (15.97%), 'lip' – 'edge' (15.83%), 'bed' – 'furniture' (15.38%), 'table' – 'furniture' (14.91%),' estuary' – 'opening' (12.77%), 'window' – 'hole/opening' (12.38%), 'beak' – 'end/point' (11.93%), 'mouth' – 'edge/tip' (10.42%), 'car' – 'cart/carriage' (10.2%), 'bark' – 'cover' (9.42%), 'tail' – 'end' (8.51%)

(v) "Bold" metaphors: 'Milky Way' – 'Trail/Road/Street' (37.5%), 'eyeball' – 'seed' (28.3%), 'estuary' – 'mouth' (25.53%), 'horizon' – 'edge/border/fringe' (24.39%), 'flame' – 'tongue' (22.02%), 'eyeball' – 'egg' (20.75%), 'sunrise' – 'come out/go out/emerge' (15.58%), 'headland' – 'nose' (15.38%), 'eclipse' – 'die/kill' (15%), 'egg' – 'testicle' (13.29%), 'bay' – 'corner' (12.77%), 'pupil' – 'child/son/daughter' (12.68%), 'semen' – 'seed' (12.12%), 'eyeball' – 'child' (11.32%), 'pupil' – 'seed/grain' (11.27%), 'testicle' – 'seed' (10.48%), 'headland'

– ‘head/forehead’ (10.26%), ‘spring’ – ‘eye’ (10.08%), ‘eclipse’ – ‘eat’ (10%), ‘uvula’ – ‘child/son’ (10%), ‘horizon’ – ‘end/finish’ (9.76%), ‘sunset’ – ‘sink’ (9.76%), ‘sunset’ – ‘enter’ (9.76%), ‘sunset’ – ‘fall/drop’ (9.76%), ‘womb’ – ‘house’ (9.57%), ‘eyeball’ – ‘fruit’ (9.43%), ‘vein’ – ‘way/street’ (9.24%), ‘sunset’ – ‘descend/go down/lower’ (8.54%), ‘meteoroid’ – ‘faeces’ (8.45%), ‘meteoroid’ – ‘tail’ (8.45%), ‘meteoroid’ – ‘fall’ (8.45%), ‘meteoroid’ – ‘fly’ (8.45%), ‘street’ – ‘manner/method/system’ (8.4%), ‘airplane’ – ‘bird’ (8.33%), ‘rainbow’ – ‘snake’ (8.27%), ‘sunrise’ – ‘appear’ (7.79%), ‘mouth’ – ‘door/entrance’ (7.64%), ‘eyelid’ – ‘lid/cover’ (7.62%), ‘eyeball’ – ‘grain’ (7.55%), ‘rainbow’ – ‘bow/arc/bend’ (7.55%), ‘spark’ – ‘lightning’ (7.5%), ‘whirlpool’ – ‘whirlwind’ (7.41%), ‘horizon’ – ‘meet/meeting’ (7.32%), ‘horizon’ – ‘basis’ (7.32%), ‘Milky Way’ – ‘river’ (7.14%)

(ix) Other: ‘lagoon’ – ‘lake/pond’ (44.64%), ‘moon’ – ‘sun’ (13.51%), ‘shadow’ – ‘photograph’ (10.95%), ‘puddle’ – ‘swamp’ (10.29%), ‘mirror’ – ‘shadow’ (8.77%), ‘wrinkle’ – ‘crease/fold/pleat’ (8.45%), ‘flood’ – ‘torrent’ (8.33%), ‘dust’ – ‘sand’ (8.09%), ‘lightning’ – ‘gleam/lighten’ (7.97%)

Thus, common metaphorical transfers, few as they are when compared with common associations by contiguity, mostly either have an additional component of biological analogy (i),[17] or are transfers that take place within the same broader semantic domain as in groups (ii), (iii), (iv), or likely have mythological or cultural underpinnings (vi). As for the associations in (vii), one could either say that they generalize the reference of a concrete referent to more abstract referents (which would be an account in terms of grundbedeutung in line with the well-known account of metaphor as making accessible more abstract domains by way of more concrete ones), or one could say that general terms may be narrowed down to more concrete referents (an account in terms of gesamtbedeutung).

These are, of course, therefore not any less metaphorical than those metaphors that cross domains, but are likely to be, from a cognitive point of view, preferred because they may be easier to process and there is no danger of communicative failure given the anchoring within the same domain (which also accounts for the fact that many of these, unlike those in group (v), occur without a contiguity anchor). The human body is likely a coherent semantic domain in psychological terms: Neely (1977) finds priming effects in a lexical decision task for parts of the body with *body* itself as prime using a short SOA (stimulus-onset-asymmetry, that is, the time elapsing between presentation of the prime and the target) even when subjects had been told before to expect mostly targets unre-

[17] Note that within this group, there are many overlaps. For instance, the association between ‘eyelid’ and ‘leather, hide’ is due to complex terms for ‘eyelid’ one of the constituents of which means ‘skin,’ but also ‘leather, hide’ (compare Appendix E, 113). Thus, in effect, the number of associations in this group may be conceived of as being smaller than reported above.

lated to the prime,[18] and priming effects are also observable for the domains of artifacts as well as living things (Moss et al. 1995, McRae and Boisvert 1998).

Camac and Glucksberg (1984), furthermore, report that there are no priming effects for pairs of words metaphorically related to each other when compared with random pairings, while there are such effects for associatively related words when compared with random pairings. Their conclusion from this is that metaphor does not operate by exploiting already existing associations, but rather is a means to create new ones.

In the light of this the evidence for common associations by metaphor can be reviewed: how would one, apart from colexification with 'star' by configurational contiguity, call the 'Milky Way' if not by some metaphorical transfer? Most languages do not feature a general term for longish accumulations of smaller entities. What would a contiguity-like conceptualization look like for meanings like 'eyeball,' apart from simple colexification with 'eye'? Most languages do not have abstract general terms for 'small round object' that does not at the same time also denote a particular small object (though Oneida and Rotokas, for instance, do). And if 'round object' is available, it is likely to have come into being through gradual extension of terms for referents which in fact are round objects (such as 'seed,' prototypically) to further referents of similar shape and size, until the term is so bleached that it does not make sense to provide an extensional definition anymore (as is evidenced by the lexical sources for classifiers with this meaning).

Brown (1999: 50-51) provides an account for generalizations as to the common semantic associations revealed in terms for items of acculturation in languages of Native America summarized in Brown (1999: 45, table 4.1.; there are both items that would be called metonymy- and metaphor-driven in the present framework on this list) by proposing a so-called "rich cognition model" sharing some aspects with Chomsky (1975), who proposed that there are innate information processing mechanisms forming a part of the language faculty:

> [W]hen different human groups are faced with the problem of giving a name to the same newly encountered object or concept, information-processing mechanisms shared by all humans are utilized to accomplish the task. This involves, at least in part, analysis of both the item to be named and of sources from which a label for the item might be retrieved. Since information-processing devices are panhuman, similar, if not the same, analyses will tend to be made, resulting in similar names for the item in question especially if sources for labels are similar.

Unfortunately, this account is only informative insofar as one presupposes knowledge about what these (not unreasonably) alleged panhuman "information-processing mechanisms" are like qualitatively. At least for the common metaphor-based denominations just mentioned, then, no grand yet unspecified theories on processing mechanisms à la Brown

[18] However, in this particular experimental setting, with longer SOA, the obtained result is the other way around, presumably because subjects require time to override the "hard-wired" associations triggered between the prime 'body' and its category members (see also Neely 1991: 285 for review).

are needed. Such theories are surely interesting, but if an explanation on a lower, less spectacular, level that takes into account established findings from psychology is available, it is to be preferred by Ockam's razor (and in fact, quite similar remarks to that in the above paragraph are found in Brown and Witkowski 1981: 606-607).

In spite of the perhaps limited contribution (typological) linguistics can make to issues concerning the workings of the human mind (see Sandra 1998), a tentative generalization would be that WHEN CONTIGUITY-BASED ASSOCIATIONS ARE READILY AVAILABLE, THEY ARE CROSS-LINGUISTICALLY PREFERRED (though "cross-linguistically preferred" does not mean that individual languages cannot not have a metaphor-driven term instead: they can). Further, when there are few possible associations by contiguity, either because there are few contiguously related entities in the real world, or because lexical counterparts for these are lacking, and IF METAPHOR HAS TO BE RESORTED TO, A WITHIN-DOMAIN TRANSFER IS PREFERRED, and only if this is unavailable as well, "bold" domain-transgressing metaphors are recurred to. This would, for instance, explain why the association between 'vein' and 'tendon,' operating within the same semantic domain, is much more frequent than that of either with 'root,' for which a crossing of domain boundaries is necessary. This, as alluded to above, is fully in line with the prevailing view of metaphor as making available less palpable semantic domains by way of more basic and familiar ones. As noted in § 3.6.2.5., however, just what a semantic domain actually is is not as straightforward as it may seem. The above discussion should not least for this reason be read as an assembly of informal thoughts on the results rather than as a full-blown theory deriving from them.

6.6. CHAPTER SUMMARY

This chapter dealt with the semantic side of lexical motivation. It provided summarizing accounts of associations within certain semantic fields, as well as discussion of common extensions of the body-part terms 'eye,' 'mouth,' 'faeces,' and also kinship terminology, to other not closely related referents. The chapter also sought for sources of non-random variation, namely in associations that can be explained by the make-up of the environment in certain regions of the world, and by examining briefly whether there is a genealogical signal within language families in semantic associations, which could not be demonstrated for at least one of the language families examined. Moreover, the data were evaluated (using a preliminary methodology) with regard to areal clusterings. This suggested that areality in lexico-semantic associations, in particular pattern of colexification, is more pervasive cross-linguistically than currently acknowledged in the relevant literature, and that they provide a huge repository for areal linguistics that is presently only scarcely exploited.

Finally, brief evaluation of cross-linguistically common associations shows that common associations are more frequently contiguity- than metaphor-driven, which can be construed as a piece of evidence for the primacy of contiguity in language.

In general, what a comparison with Buck (1949) shows is that many of the cross-linguistically attested semantic associations are mirrored in individual languages of one particular family (compare cross-references in Appendix E). Still, it becomes clear that to

really be able to make solid inferences about likely associations from a cross-linguistic point of view, it is not enough to just investigate one particular family and that doing so may lead to a distorted picture of the possible cross-linguistic variation, as for instance the absence of lexical associations of 'milk' and 'breast' and 'tear' and 'water,' which are both extremely frequent world-wide, in Indo-European show (compare relevant data in Buck 1949).

Chapter 7

Prospects for Cross-Linguistic Research on the Lexicon

This study hopefully demonstrated two things: first, that the lexicon is not just merely "an appendix of the grammar, a list of basic irregularities" (Bloomfield 1933: 274), a doctrine that is still very much alive in many theoretical approaches to Linguistics, but that its formal structure is systematically determined by complexity of the roots and of the sound system. Neither are semantic structures completely random, but they are both amenable to areal influence in colexifying and analyzable terms, and, with regard to the latter, they co-vary to some extent with the type of word-formation most commonly used in individual languages.

However, this study is not the ultimate statement on comparative lexicology (as defined in the introductory chapter 1), but rather should be seen as a first attempt to probe largely uncharted terrain, at least from an explicitly cross-linguistic point of view (and this becomes especially, but not only, clear from the fact that it is restricted to a mere 160 meanings, while the lexicon in reality is of course much richer, a vast repository of linguistic and culture knowledge).

It does not discuss all possible matters, and surely many more interactions between grammar and lexicon as well as systematic tendencies in the lexicon are to be discovered in the future.

First, there are limitations on what can be achieved by extracting data from dictionaries. What is required are in-depth fieldwork-based studies to establish cross-linguistic variation. Heath and McPherson (2009) are exemplary in combining an in-depth investigation of individual languages with generalization across languages, thus bridging the divide between too narrow of a focus on phenomena in a single language (although, surely, also these can be interesting), and a necessarily more coarse-grained typological study based on extant materials. Such a marriage, though labour-intensive, is needed to be able to avoid, or at least to mitigate, some of the pitfalls described in chapter 3 that come along with working on extant sources.

But even within the framework of broad typological investigations, not all potentially relevant factors were in fact addressed in the present study. For one, an aspect of interaction between lexicon and grammar that has received unduly little attention are grammatical aspects that may render terms for, e.g. a body-part, less necessary. For instance, the Nuuchahnulth word for 'finger' *ċaċałaqṅukum* is clearly morphologically complex, containing the root *ċałaq-* and *-ṅukum* 'in or at the hand.' Now, Nuuchahnulth has a quite frequent prefix *ƙum-* 'point, poke, press with finger' (Davidson 2002: 63) that is prefixed to a verb when the action is carried out with the fingers, thus reducing the need to employ the morphologically complex term. For instance, *ƙum-'aqƛ* is 'have one's finger poked inside' (Davidson 2002: 64, *-'aqƛ* 'inside').

Another area in which it falls short is the investigation of specific patterns of colexification and possible correlations with structural properties of languages as claimed to exist by Klimov (1977) and Lehmann (2002) for active-stative languages. These clearly deserve further investigation under the perspective of interaction between grammar and lexicon (though this merely reiterates Nichols's 1992: 260 request). Somewhat similarly, Nichols (2008: 684) suggestion for "a new kind of typological classification of languages according to noun root lexical properties" is a promising avenue for further research.

This study paid attention to the nominal lexicon specifically, and certain findings are indeed restricted to referring expressions. As pointed out by Talmy (2000: 59endnote11), there is a complementary perspective on the verbal lexicon that would in principle be at least equally interesting to investigate.

Furthermore, it seems promising to investigate if research in Social Psychology discussed in § 5.7.6. on different kinds of reasoning and their prevalence in different parts of the world has effects on the semantic structuring of the lexicon, once a clearer picture of the distribution of that prevalence emerges.

In summary, cross-linguistic investigation of structures in the lexicon is a field wide open for new discoveries to be made. Since words, of course, represent concepts and are thus intimately linked with cognitive representations, such research has the potential to strengthen the interdisciplinary links between linguistics, in particular with a cross-linguistic orientation, and neighboring fields of research such as Cognitive and Social Psychology, and to engage the disciplines in a productive dialogue. The questions are in principle all open to empirical investigation.

Appendix A:
Sample Languages, Areal Affiliation, and Consulted Sources

Language	Sample	Classification (from Dryer 2005)	Area Dryer-6	Area Nichols-11	Area Nichols-3	Consulted Source(s)
STATISTIC						
Hausa	Statistic	Afro-Asiatic, Chadic, West Chadic	Africa	Africa	Old World	Bargery and Westermann 1934, Newman 2000
Katcha	Statistic	Kadugli	Africa	Africa	Old World	Stevenson n.d., Tucker et al. 1966
Khoekhoe	Statistic	Khoisan, Central Khoisan	Africa	Africa	Old World	Haacke and Eiseb 2002, Hagman 1977
Mbum	Statistic	Niger-Congo, Adamawa-Ubangi, Adamawa	Africa	Africa	Old World	Hino 1978, Hagege 1970
Ngambay	Statistic	Nilo-Saharan, Central Sudanic, Bongo Bagirmi	Africa	Africa	Old World	Moser and Dingatoloum 2001, Djarangar 1989, Thayer 1978
Baruya	Statistic	Trans-New-Guinea,Angan	Australia-New Guinea	New Guinea	Pacific	Lloyd 1992, Lloyd and Healey 1970
Berik	Statistic	Tor	Australia-New Guinea	New Guinea	Pacific	Westrum and Westrum 1986
Buin	Statistic	East Bougainville	Australia-New Guinea	Oceania	Pacific	Laycock 2003, Anonymous n.d.

App. A (cont'd)

Kaluli	Statistic	Bosavi	Australia-New Guinea	New Guinea	Pacific	Schieffelin and Feld 1998, Grosh and Grosh 2000
Kwoma	Statistic	Middle Sepik	Australia-New Guinea	New Guinea	Pacific	Bowden 1997
Mali	Statistic	Baining-Taulil	Australia-New Guinea	Oceania	Pacific	Stebbins with Tayul in press
Meyah	Statistic	East Bird's Head	Australia-New Guinea	New Guinea	Pacific	Gravelle 2004
Rotokas	Statistic	West Bougainville, North Halmaheran	Australia-New Guinea	Oceania	Pacific	Firchow 1970, Firchow and Firchow 2008, Robinson 2011
Sahu	Statistic	West Papuan, West Papuan	Australia-New Guinea	South & Southeast Asia	Pacific	Visser and Voorhoeve 1987
Toaripi	Statistic	Eleman	Australia-New Guinea	New Guinea	Pacific	Brown 1968, 1972
Yir Yoront	Statistic	Australian, Pama-Nyungan	Australia-New Guinea	Australia	Pacific	Alpher 1991
Abzakh Adyghe	Statistic	Northwest Caucasian	Eurasia	Europe	Old World	Paris and Batouka 2005, Hewitt 2005
Badaga	Statistic	Dravidian, Southern Dravidian	Eurasia	South & Southeast Asia	Old World	Hockings and Pilot-Raichoor 1992, Balakrishnan 1999
Basque	Statistic	Basque	Eurasia	Europe	Old World	Morris 1998, Trask 1997

App. A (cont'd)

Bezhta	Statistic	Nakh-Daghestanian, Daghestanian, Avar-Andic-Tsezic	Eurasia	Europe	Old World	Comrie and Khalilov 2009a, b, Zaira Khalilova p.c.
Chukchi	Statistic	Chukotko-Kamtchatkan	Eurasia	Eurasia	Old World	Fortescue 2005, supplemented by Kurebito 2001, Dunn 1999
Ket	Statistic	Yeniseian	Eurasia	Eurasia	Old World	Vajda and Nefedov 2009, Andrej Nefedov p.c., Vajda 2004b
Khalkha[1]	Statistic	Altaic, Mongolic	Eurasia	Eurasia	Old World	Lessing 1995, Poppe 1954, Street 1963
Laz	Statistic	Kartvelian	Eurasia	Europe	Old World	Johanna Mattissen with Sevim Genç p.c., Anderson 1963
Nivkh	Statistic	Nivkh	Eurasia	Eurasia	Old World	Saveleva and Taksami 1965, 1970, Gruzdeva 1998, Ekaterina Gruzdeva p.c.
Kildin Saami	Statistic	Uralic, Finnic	Eurasia	Europe	Old World	Rießler 2009
Welsh	Statistic	Indo-European, Celtic	Eurasia	Europe	Old World	Evans and Thomas 1983, Williams 1980, Thorne 1993
Kolyma Yukaghir	Statistic	Yukaghir	Eurasia	Eurasia	Old World	Endo 1997, 2001, Nikolaeva 2006, Maslova 1998

1 Khalkha is in fact one of the major varieties of spoken Mongolian, while the sources from which data for this study stem actually describe the written language, which differs in some respects from the spoken language phonologically and grammatically (the written language does not have a WALS code of its own).

App. A (cont'd)

Biloxi	Statistic	Siouan	North America	Eastern North America	New World	Dorsey and Swanton 1912, Einaudi 1974
Carrier	Statistic	Na-Dene, Athabascan	North America	Western North America	New World	Morice 1932, Story 1984, Gessner 2003
Upper Chehalis	Statistic	Salishan, Tsamosan	North America	Western North America	New World	Kinkade 1963, 1991
Cheyenne	Statistic	Algic, Algonquian	North America	Eastern North America	New World	Fisher et al. 2008
Chickasaw	Statistic	Muskogean, Western	North America	Eastern North America	New World	Munro and Willmond 1994 and p.c.
Highland Chontal	Statistic	Tequistlatecan	North America	Mesoamerica	New World	Turner 1966, Turner and Turner 1971
Ineseño Chumash	Statistic	Chumash	North America	Western North America	New World	Applegate 1972, Santa Ynez Band of Chumash Indians 2008
Haida	Statistic	Haida	North America	Western North America	New World	Enrico 2005, Levine 1977
San Mateo del Mar Huave	Statistic	Huavean	North America	Mesoamerica	New World	Stairs Kreger and de Stairs 1981, Archivo de Lenguas Indígenas de México 1983
Itzaj	Statistic	Mayan	North America	Mesoamerica	New World	Hofling and Tesucún 1997, 2000
Kiliwa	Statistic	Hokan, Yuman	North America	Western North America	New World	Mixco 1965, 1985, 2000
Kiowa	Statistic	Kiowa-Tanoan	North America	Eastern North America	New World	Harrington 1928, Watkins 1984

App. A (cont'd)

Nez Perce	Statistic	Penutian, Sahaptian	North America	Western North America	New World	Aoki 1970, 1994, Rude 1985
Nuuchahnulth	Statistic	Wakashan, Northern Wakashan	North America	Western North America	New World	Nakayama 2001, Stonham 2005, Davidson 2002
Oneida	Statistic	Iroquoian, Northern Iroquian	North America	Eastern North America	New World	Abbott 2000, Michelson and Doxtator 2002
Santiago Mexquititlan Otomí	Statistic	Oto-Manguean, Otomian	North America	Mesoamerica	New World	Bakker and Hekking 2009. Arroyo 1955
Pawnee	Statistic	Caddoan	North America	Eastern North America	New World	Parks 1976, Parks and Pratt 2008
Pipil	Statistic	Uto-Aztecan, Aztecan	North America	Mesoamerica	New World	Campbell 1985
Xicotepec de Juárez Totonac	Statistic	Totonacan	North America	Mesoamerica	New World	Reid and Bishop 1974, Reid and Watson 1991
Wappo	Statistic	Wappo-Yukian, Wappo	North America	Western North America	New World	Sawyer 1965, Thompson et al. 2006
Central Yup'ik	Statistic	Eskimo-Aleut	North America	NA	New World	Jacobson 1984, Mather et al. 2002
Copainalá Zoque	Statistic	Mixe-Zoque	North America	Mesoamerica	New World	Harrison et al. 1981
Aguaruna	Statistic	Jivaroan	South America	South America	New World	Wipio Deicat 1996 supplemented by Corbera Mori 1994, Overall 2007
Arabela	Statistic	Zaparoan	South America	South America	New World	Rich 1999

App. A (cont'd)

Aymara	Statistic	Aymaran	South America	South America	New World	Cotari 1978, Gómez Bacarreza and Condori Cosme 1992, Hardman 2001
Bora	Statistic	Huitotoan	South America	South America	New World	Thiesen and Thiesen 2008, Thiesen 1996, Seifart 2005 and p.c.
Bororo	Statistic	Macro-Ge, Bororo	South America	South America	New World	Ochoa Camargo1997, Crowell 1979, Canzio 1997
Carib	Statistic	Carib	South America	South America	New World	Renault-Lescure 2009 supplemented by Courtz 2008
Cashinahua	Statistic	Panoan	South America	South America	New World	Montag 2008, Camargo 1991
Cavineña	Statistic	Tacanan	South America	South America	New World	Camp and Liccardi 1989 supplemented by Guillaume 2008
Cayapa	Statistic	Barbacoan	South America	South America	New World	Lindskoog and Lindskoog 1964, Vittadello 1988
Chayahuita	Statistic	Cahuapanan	South America	South America	New World	Hart 2008, Wise 1999, Hart et al. 1976
Cubeo	Statistic	Tucanoan	South America	South America	New World	Morse et al. 1999, Morse and Maxwell 1999
Embera	Statistic	Choco	South America	South America	New World	Sara 2001, Aguirre Licht 1999

App. A (cont'd)

Guaraní	Statistic	Tupian, Tupi-Guaraní	South America	South America	New World	Guasch and Ortiz 1998, supplemented by Britton 2005
Hupda	Statistic	Vaupés-Japurá	South America	South America	New World	Epps 2008, 2009
Jarawara	Statistic	Arauan	South America	South America	New World	Vogel 2006, Dixon and Vogel 2004
Miskito	Statistic	Misumalpan	South America	South America	New World	Warman 1959, Heath 1950, Heath and Marx 1961, Salamanca 1988, Adam 1968
Piro	Statistic	Arawakan	South America	South America	New World	Matteson 1965, supplemented by Nies 1986
Imbabura Quechua	Statistic	Quechua	South America	South America	New World	Gómez Rendón 2009, Cole 1982, Adelaar with Muysken 2004
Rama	Statistic	Chibchan	South America	South America	New World	Rigby and Schneider 1989, Craig 1990
Wichí	Statistic	Matacoan	South America	South America	New World	Viñas Urquiza 1974, Vidal and Nercesian 2009
Yanomámi	Statistic	Yanomam	South America	South America	New World	Lizot 1975, Ramirez 1994, Mattéi-Muller 2009
Bislama	Statistic	Pidgins & Creoles, English-Based	Southeast Asia & Oceania	Oceania	Pacific	Crowley 2003, 2004
Great Andamanese	Statistic	Andamanese	Southeast Asia & Oceania	South & Southeast Asia	Old World	Man 1923, Manoharan 1989, Abbi 2006

App. A (cont'd)

Bwe Karen	Statistic	Sino-Tibetan, Tibeto-Burman, Karen	Southeast Asia & Oceania	South & Southeast Asia	Old World	Henderson 1997
White Hmong	Statistic	Hmong-Mien	Southeast Asia & Oceania	South & Southeast Asia	Old World	Mottin 1978, Ratliff 1992, 2009
Sedang	Statistic	Austro-Asiatic, Mon-Khmer, Bahnaric	Southeast Asia & Oceania	South & Southeast Asia	Old World	Smith 1979, 2000
Tetun	Statistic	Austronesian, Central Malayo-Polynesian	Southeast Asia & Oceania	South & Southeast Asia	Pacific	Hull 2001, Hull and Eccles 2004, Williams-Van Klinken et al. 2002, van Engelenhoven and Williams-Van Klinken 2005
Yay	Statistic	Tai-Kadai, Kam-Tai	Southeast Asia & Oceania	South & Southeast Asia	Old World	Hudak 1991
CORE						
Buli	Core	Niger-Congo, Gur	Africa	Africa	Old World	Kröger 1992
Efik	Core	Niger-Congo, Benue-Congo, Cross-River	Africa	Africa	Old World	Goldie 1964, Welmers 1968
Kanuri	Core	Nilo-Saharan, Western Saharan	Africa	Africa	Old World	Löhr et al. 2009, Hutchison 1981
Dongolese Nubian	Core	Nilo-Saharan, Eastern Sudanic, Nubian	Africa	Africa	Old World	Armbruster 1960, 1965
Rendille	Core	Afro-Asiatic, Cushitic, Eastern Cushitic	Africa	Africa	Old World	Pillinger 1989, Pillinger and Galboran 1999
Burarra	Core	Australian, Burarran	Australia-New Guinea	Australia	Pacific	Glasgow 1994, Green 1987

App. A (cont'd)

Kyaka	Core	Trans-New-Guinea, Engan	Australia-New Guinea	New Guinea	Pacific	Draper and Draper 2002
Nunggubuyu	Core	Australian, Guwinyguan, Nunggubuyu	Australia-New Guinea	Australia	Pacific	Heath 1982, 1984
Kosarek Yale	Core	Trans-New-Guinea, Mek	Australia-New Guinea	New Guinea	Pacific	Heeschen 1992
Greek	Core	Indo-European, Greek	Eurasia	Europe	Old World	Pring 1982, Joesph and Philippaki-Warburton 1987
Sora	Core	Austro-Asiatic, Munda	Eurasia	South & Southeast Asia	Old World	Ramamurti 1938, Anderson and Harrison 2008
Wintu	Core	Penutian, Wintuan	North America	Western North America	New World	Pitkin 1984, 1985
Hawaiian	Core	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Oceania	Pukui and Elbert 1987, Elbert 1979
Manange	Core	Sino-Tibetan, Tibeto-Burman, Bodic	Southeast Asia & Oceania	South & Southeast Asia	Old World	Hildebrandt 2003, 2009
Mandarin	Core	Sino-Tibetan, Chinese	Southeast Asia & Oceania	South & Southeast Asia	Old World	Lin 2001, Wiebusch 2009
Vietnamese	Core	Austro-Asiatic, Mon-Khmer, Viet-Muong	Southeast Asia & Oceania	South & Southeast Asia	Old World	Alves 2009, Thompson 1965
EXT-1						
Anggor	Ext-1	Sengai	Australia-New Guinea	New Guinea	Pacific	Litteral n.d., Litteral 1980
Dadibi	Ext-1	Teberan-Pawaian, Teberan	Australia-New Guinea	New Guinea	Pacific	Bai and Whitby 2006

App. A (cont'd)

Kemtuik	Ext-1	Nimboran	Australia-New Guinea	New Guinea	Pacific	Kroneman and Duha 2005
Lavukaleve	Ext-1	Solomons-East Papuan	Australia-New Guinea	Oceania	Pacific	Terrill 1999, Terrill 1996-2003
One	Ext-1	Torricelli, West Papei	Australia-New Guinea	New Guinea	Pacific	Donohue n.d. a.
Rao	Ext-1	Lower Sepik-Ramu, Annaberg	Australia-New Guinea	New Guinea	Pacific	Stanhope 1980
Sentani	Ext-1	Sentani	Australia-New Guinea	New Guinea	Pacific	Cowan 1965
Sko	Ext-1	Sko, Western Sko	Australia-New Guinea	New Guinea	Pacific	Kemo et al. 2002
Tasmanian[2]	Ext-1	Tasmanian	Australia-New Guinea	Australia	Pacific	Schmidt 1952, Plomley 1968
Waris	Ext-1	Border	Australia-New Guinea	New Guinea	Pacific	Brown and Walsa Translation Team 2007
Yei	Ext-1	Morehead and Upper Maro Rivers	Australia-New Guinea	New Guinea	Pacific	Donohue n.d. b.
Acoma	Ext-1	Keresan	North America	Eastern North America	New World	Miller 1965, supplemented by Maring 1967
Quileute	Ext-1	Chimakuan	North America	Western North America	New World	Andrade 1933, Powell and Woodruff 1976

[2] Only terms considered to be safely attested by Schmidt (1952) and those mentioned in the thematic glossary (Schmidt 1952: 469-516) are considered.

App. A (cont'd)

Abipón	Ext-1	Guaicuruan	South America	South America	New World	Najlis 1966
Lengua	Ext-1	Mascoian	South America	South America	New World	Lowes 1954, Perasso and Bracco 1979
Macaguán	Ext-1	Guahiban	South America	South America	New World	Buenaventura 1993
Sáliba	Ext-1	Sáliban	South America	South America	New World	Benaissa 1991, supplemented by Suárez 1977
Tehuelche	Ext-1	Chon, Chon Proper	South America	South America	New World	Fernández Garay 1998, 2004
Toba	Ext-1	Guaicuruan	South America	South America	New World	Buckwalter 1980, Messineo 2003
EXT-2						
Bakueri	Ext-2	Niger-Congo, Benue-Congo, Bantoid	Africa	Africa	Old World	Kagaya 1992
Koyraboro Senni	Ext-2	Nilo-Saharan, Songhay	Africa	Africa	Old World	Heath 1998, 1999
Noni	Ext-2	Niger-Congo, Benue-Congo, Bantoid	Africa	Africa	Old World	Lux 2003, Hyman 1981
Swahili	Ext-2	Niger-Congo, Benue-Congo, Bantoid	Africa	Africa	Old World	Schadeberg 2009
Yoruba	Ext-2	Niger-Congo, Benue-Congo, Defoid	Africa	Africa	Old World	University Press PLC n.d., Joseph Atoyebi p.c.
Gurindji	Ext-2	Australian, Pama-Nyungan	Australia-New Guinea	Australia	Pacific	McConvell 2009
Muna	Ext-2	Austronesian, Western Malayo-Polynesian, Sulawesi	Southeast Asia & Oceania	South & Southeast Asia	Pacific	Van den Berg and Sidu 1996

App. A (cont'd)

Ngaanyatjarra	Ext-2	Australian, Pama-Nyungan	Australia-New Guinea	Australia	Pacific	Glass and Hackett 2003
Japanese	Ext-2	Japanese	Eurasia	Eurasia	Old World	Schmidt 2009
Blackfoot	Ext-2	Algic, Algonquian	North America	Eastern North America	New World	Frantz 1991, Frantz and Russell 1995
Cahuilla	Ext-2	Uto-Aztecan, Taki	North America	Western North America	New World	Seiler 1977, Seiler and Hioki 1979
Comanche	Ext-2	Uto-Aztecan, Numic	North America	Eastern North America	New World	Robinson and Armagost 1994, Charney 1993
Kashaya	Ext-2	Hokan, Pomoan	North America	Western North America	New World	Oswalt 1961, 1975, n.d., Cengerova et al. 2009
Lake Miwok	Ext-2	Penutian, Utian	North America	Western North America	New World	Callaghan 1963, 1965
Lakhota	Ext-2	Siouan	North America	Eastern North America	New World	Rood 1976, Ingham 2003
Lesser Antillean Creole French	Ext-2	Pidgins & Creoles, French-Based	North America	Eastern North America	New World	Mondesir 1992, Carrington 1984
Tuscarora	Ext-2	Iroquoian, Northern Iroquoian	North America	Eastern North America	New World	Rudes 1999
Yana	Ext-2	Hokan, Yana	North America	Western North America	New World	Sapir and Swadesh 1960
Yaqui	Ext-2	Uto-Aztecan, Cahita	North America	Mesoamerica	New World	Estrada Fernández 2009a, b
Yuki	Ext-2	Wappo-Yukian, Yukian	North America	Western North America	New World	Sawyer and Schlichter 1984

App. A (cont'd)

San Lucas Quiaviní Zapotec	Ext-2	Oto-Manguean, Zapotecan	North America	Mesoamerica	New World	Munro and Lopez 1999, Lee 1999
Huambisa	Ext-2	Jivaroan	South America	South America	New World	Jakway et al. 1987
Kaingang	Ext-2	Macro-Ge, Ge-Kaingang	South America	South America	New World	Wiesemann 1972, Wiesemann Gojtéj 2011
Maxakalí	Ext-2	Macro-Ge, Maxakalí	South America	South America	New World	Popovich and Popovich 2005
Ancash Quechua	Ext-2	Quechua	South America	South America	New World	Carranza and Lustig 2003, supplemented by Parker and Chavez 1976, Adelaar with Muysken 2004
Tsafiki	Ext-2	Barbacoan	South America	South America	New World	Moore 1966, Dickinson 2003
Wayampi	Ext-2	Tupian, Tupi-Guaraní	South America	South America	New World	Grenand 1980, 1989
Fijian	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Schütz 1985, Capell 1991
Hani	Ext-2	Sino-Tibetan, Tibeto-Burman, Burmese-Lolo	Southeast Asia & Oceania	South & Southeast Asia	Pacific	Lewis and Bibo 1997
Kapingamarangi	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Lieber and Dikepa 1974
Lenakel	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Lynch 1977, 1978

App. A (cont'd)

Malagasy	Ext-2	Austronesian, Western Malayo-Polynesian, Borneo	Southeast Asia & Oceania	Africa	Pacific	Dez 1980, Adelaar 2009
Rotuman	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Inia et al.1998, Churchward 1998/1940, Vamarasi 2002
Samoan	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Milner 1993, Mosel and Hovdhaugen 1992
Takia	Ext-2	Austronesian, Eastern Malayo-Polynesian, Oceanic	Southeast Asia & Oceania	Oceania	Pacific	Ross 2002, 2009

Appendix B: Percentages of analyzable terms and the contiguity-similarity ratio

	Analyzability All Domains	**Thereof Derived %**	Modified Simplicity Score Haspelmath and Tadmor	Nature	Artifacts	Body-Parts	Phases of the day & Misc.	**Colexification All Domains**	**Contiguity-Similarity Ratio**	Analyzable only	Colexifying only	Nature	Artifacts	Body-Parts	Phases of the day & Misc.
CORE															
Hausa	**13.4**	**29.5**	24.18*	2.7	2.5	6.0	2.2	**33.8**	**1.27**	1.09	2.36	2.09	0.7	0.93	1.09
Katcha	**15.9**	**0**	NA	5.4	1.8	7.9	0.9	**11.6**	**0.69**	1.02	0.53	0.75	0.14	0.89	1.0
Khoekhoe	**31.6**	**36.3**	NA	12.9	6.3	9.0	3.3	**18.7**	**0.64**	0.36	3.42	0.76	0.39	0.68	0.4
Mbum	**29.3**	**2.7**	NA	9.8	4.5	13.0	2.0	**10.2**	**1.15**	0.93	0.53	1.47	0.6	1.89	0
Ngambay	**19.7**	**0**	NA	4.2	3.9	6.5	5.2	**38.9**	**0.82**	0.51	1.68	1.21	0.35	0.73	0.4
Baruya	**19.6**	**0**	NA	6.7	4.8	6.9	1.2	**30.7**	**0.75**	0.66	2.17	1.51	0.31	0.69	0
Berik	**15.3**	**0**	NA	7.3	0	7.6	0.4	**15.3**	**0.78**	0.45	1.11	0.5	2.0	2.0	0
Buin	**6.4**	**48.9**	NA	4.5	0	1.1	0.8	**37.8**	**0.64**	0.3	1.53	1.02	0	0.55	0
Kaluli	**19.4**	**0**	NA	9.7	4.9	1.9	2.9	**18.9**	**1.1**	0.38	2.9	2.11	2.5	0.29	0
Kwoma	**20.9**	**0**	NA	6.0	2.9	10.5	1.5	**40.7**	**1.31**	1.34	2.94	1.55	0.75	2.15	0.18
Mali	**9.3**	**41.7**	NA	2.3	2.8	3.7	0.5	**20.1**	**0.90**	0.36	1.43	1.32	1.33	0.27	NA
Meyah	**25.2**	**0**	NA	10.3	4.7	5.6	4.7	**12.9**	**1.43**	1.21	2.46	1.04	1.5	2.0	3.0
Rotokas	**30.5**	**52.1**	NA	9.5	11.4	5.4	4.1	**22.6**	**0.55**	0.21	1.48	1.03	0.42	0.46	0.22
Sahu	**13.6**	**30.3**	NA	7.0	0	5.8	0.9	**26.5**	**1.27**	1.49	1.86	1.54	NA	1.15	0
Toaripi	**27.2**	**0**	NA	7.8	4.8	11.3	3.2	**36.4**	**1.20**	0.68	2.61	1.54	0.6	1.23	1.0
Yir Yoront	**21.4**	**0.6**	NA	6.5	6.7	4.0	4.2	**37.1**	**1**	0.78	1.7	1.4	0.91	0.67	1.0

App. B (cont'd)

Abzakh Adyghe	**37.4**	**2.4**	NA	13.6	5.7	15.8	2.4	**24.3**	**1.11**	0.73	2.27	1.5	1.5	0.69	0.63
Badaga	**9.0**	**0**	NA	0.7	0.7	5.4	2.3	**38.4**	**0.75**	1.08	1.43	0.72	1.13	1.04	0.16
Basque	**18.0**	**11.8**	NA	7.6	1.5	6.0	2.9	**42.0**	**1.83**	1.13	3.98	2.53	1.44	1.73	0.12
Bezhta	**6.9**	**10**	30.96	3.1	2.3	0.8	0.8	**21.7**	**1.01**	0.33	2.52	1.23	0.4	2.5	0
Chukchi	**13.1**	**50**	NA	6.1	4.4	0.8	1.7	**15.0**	**0.47**	0.27	2.0	0.51	0.46	0.89	0
Ket	**30.5**	**1.9**	40.68	15.1	4.5	8.8	2.1	**18.5**	**0.78**	0.42	2.75	1.46	0.56	0.5	0
Khalkha	**12.9**	**33.9**	NA	6.4	1.8	2.7	2.0	**47.9**	**1.38**	1.21	2.67	1.21	2.01	1.81	0.37
Laz	**7.5**	**23.5**	NA	3.9	1.8	1.8	0	**11.5**	**0.51**	0.6	0.72	1.09	0	0.4	NA
Nivkh	**14.7**	**0**	NA	7.6	1.6	2.8	2.8	**17.2**	**0.58**	0.65	1.0	0.55	1.17	0.57	0.33
Kildin Saami	**7.4**	**0**	25.08	3.1	0.8	2.7	0.8	**16.4**	**0.92**	1.11	1.8	2.13	0.4	0.43	0
Welsh	**15.1**	**20.1**	NA	7.2	1.6	4.4	1.9	**20.4**	**0.68**	0.81	1.43	0.47	0	1.32	0.4
Kolyma Yukaghir	**24.7**	**17.6**	NA	7.1	2.7	12.9	1.8	**15.3**	**0.98**	0.57	1.77	1.1	8.0	0.61	NA
Biloxi	**32.5**	**12.8**	NA	11.7	9.2	6.7	5.0	**13.3**	**0.84**	0.67	1.46	0.88	0.32	1.25	NA
Carrier	**31.4**	**19.6**	NA	14.5	9.3	5.6	2.1	**13.5**	**1.26**	0.96	2.72	1.36	0.68	2.1	1.43
Upper Chehalis	**25.2**	**42.3**	NA	11.9	6.4	4.5	2.4	**19.4**	**0.34**	0.28	1.06	0.57	0.21	0.13	0.25
Cheyenne	**30.3**	**2.4**	NA	11.9	10.7	4.5	3.2	**26.5**	**0.48**	0.63	1.04	1.03	0.24	0.19	0.5
Chickasaw	**41.8**	**24.7**	NA	14.2	12.4	11.6	3.6	**18.3**	**0.53**	0.35	1.98	0.98	0.22	0.43	0.52
Highland Chontal	**13.3**	**0**	NA	6.6	0	5.8	0.9	**26.8**	**0.98**	1.55	1.46	2.0	0	0.57	0.18
Ineseño Chumash	**13**	**61.2**	NA	7.1	1.9	2.4	1.6	**28.2**	**0.62**	0.44	1.73	0.83	0.43	0.58	0
Haida	**19.1**	**19.4**	NA	7.0	6.3	5.6	0.2	**30.1**	**0.72**	0.33	1.9	0.85	0.19	1.44	0
San Mateo del Mar Huave	**33.1**	**16.9**	NA	17.3	0.2	13.4	2.2	**14.6**	**1.20**	0.97	1.85	1.83	0.25	0.75	2.0
Itzaj	**22.3**	**10.7**	NA	7.9	1.5	9.6	3.2	**28.5**	**1.15**	0.99	2.44	1.07	0.24	1.86	1.17
Kiliwa	**50.2**	**22.1**	NA	19.8	10.7	15.2	4.5	**12.6**	**0.89**	0.6	0.79	2.12	0.44	0.71	0.18
Kiowa	**40.8**	**23.2**	NA	13.6	10.3	14.2	2.6	**23.0**	**0.32**	0.15	1.05	0.29	0.08	0.6	0.33
Nez Perce	**21.7**	**52.8**	NA	6.4	9.0	3.5	2.8	**37.5**	**0.57**	0.32	1.92	0.96	0.3	0.43	0.5

App. B (cont'd)

Nuuchahnulth	**22.4**	**31.1**	NA	5.8	8.9	6.0	1.7	**19.9**	**0.14**	0.1	1.31	0.19	0.05	0.12	1.0
Oneida	**23.2**	**48.3**	NA	6.2	8.5	7.0	1.6	**19.4**	**0.25**	0.08	1.08	0.35	0.3	0.15	0
Santiago Mexquititlan Otomí	**24.1**	**12.0**	37.9	11.2	3.6	8.2	1.2	**13.4**	**0.84**	0.74	1.01	1.24	0.84	0.52	0
Pawnee	**46.7**	**14.2**	NA	17.5	12.2	12.4	4.5	**25.5**	**0.74**	0.53	2.7	0.55	0.56	3.75	0.2
Pipil	**18.2**	**25.9**	NA	6.0	0.9	10.4	0.9	**21.9**	**1.07**	0.66	3.21	0.71	1.2	1.5	NA
Xicotepec de Juárez Totonac	**12.0**	**59.3**	NA	5.9	3.4	1.1	1.7	**21.0**	**0.48**	0.2	0.9	0.83	0.3	1.44	0.17
Wappo	**12.9**	**0**	NA	4.3	0.9	6.0	1.7	**22.0**	**1.32**	1.0	1.55	1.75	0.6	1.17	NA
Central Yup'ik	**33.5**	**90.6**	NA	13.7	7.4	8.9	3.5	**19.0**	**0.82**	0.5	2.16	1.57	0.29	1.32	0
Copainalá Zoque	**14**	**6.7**	NA	2.8	2.8	7.5	0.9	**22.0**	**0.73**	0.63	1.94	0.46	0.5	1.75	0
Aguaruna	**12.1**	**42.0**	NA	2.4	3.9	4.3	1.5	**31.8**	**0.53**	0.4	1.34	0.8	0	0.57	1.6
Arabela	**23.6**	**29.3**	NA	6.1	5.9	10.8	0.8	**37.0**	**0.97**	0.45	2.06	1.26	0.4	1.13	0
Aymara	**4.9**	**36.4**	NA	2.2	0.9	1.8	0	**27.1**	**0.58**	1.0	0.99	0.57	0.5	0.67	0
Bora	**34.2**	**70.8**	NA	9.6	11.3	11.1	2.2	**22.3**	**0.57**	0.19	2.44	1.79	0.08	0.34	0.2
Bororo	**39.5**	**6.8**	NA	15.0	9.5	9.7	5.3	**28.8**	**0.64**	0.4	2.59	1.28	0.17	0.56	0
Carib	**17.9**	**51**	NA	5.5	2.9	7.3	2.2	**21.2**	**0.78**	0.36	2.22	2.27	0.4	0.41	0
Cashinahua	**12**	**50**	NA	1.4	8.8	1.9	0	**39.4**	**0.78**	0.24	1.83	2.11	0.29	0.82	0
Cavineña	**17.2**	**12.6**	NA	4.5	1.3	7.1	4.3	**21.9**	**1.06**	0.26	2.08	1.09	1.33	0.97	1.0
Cayapa	**20.8**	**4.4**	NA	7.4	0.9	10.6	1.9	**21.6**	**1.60**	1.81	1.8	1.33	3.0	1.78	0
Chayahuita	**31.8**	**49.1**	NA	13.9	4.1	11.8	2.1	**16.6**	**1.01**	0.85	2.49	0.98	1.8	1.06	0
Cubeo	**42.0**	**37.0**	NA	11.8	9.1	19.4	1.7	**14.0**	**0.35**	0.17	2.4	0.5	0.04	0.52	1
Embera	**28.2**	**4.4**	NA	13.3	1.1	12.1	1.7	**16.7**	**0.49**	0.38	0.96	0.69	1.27	0.42	0
Guaraní	**29.7**	**16.7**	NA	14.6	5.1	9.2	0.8	**27.2**	**1.11**	0.91	2.66	1.95	0.13	0.93	0
Hupda	**29.4**	**2.7**	37.12	11.3	6.0	10.5	1.6	**26.2**	**1.62**	0.69	4.0	2.21	1.0	1.43	2
Jarawara	**23.4**	**14**	NA	5.6	1.9	14.0	1.9	**33.2**	**0.71**	0.65	1.84	1.06	0.83	0.52	0
Miskito	**29.0**	**5.6**	NA	13.2	5.1	9.7	1.1	**30.3**	**0.62**	0.44	1.38	1.39	0.18	0.34	0
Piro	**25.0**	**7.8**	NA	7.3	7.3	9.2	1.1	**22.1**	**0.58**	0.43	1.66	0.85	0.39	0.75	0

App. B (cont'd)

Imbabura Quechua	**17.7**	**8.9**	30.96	5.5	2.4	8.3	1.6	**14.6**	**1.67**	1.44	3.62	3.13	0	1.88	0
Rama	**30.2**	**2.9**	NA	13.2	0.9	13.5	2.6	**27.1**	**1.33**	1.36	1.69	1.72	1.75	1.27	0.4
Wichí	**25**	**17.5**	33.28	8.3	5.6	10.3	0.8	**7.1**	**0.65**	0.58	0.8	1.0	0	1.0	0
Yanomami	**35.0**	**31.9**	NA	11.9	7.5	14.6	1.0	**31.0**	**0.52**	0.3	1.93	0.7	0.66	0.33	0
Great Andamanese	**22.4**	**32.4**	NA	14.4	0	8.0	0	**8.9**	**0.85**	1.04	1.19	1.31	NA	0.5	0
Bislama	**18.2**	**2.8**	NA	6.1	0.2	9.7	2.3	**36.3**	**1.31**	0.99	3.29	2.25	1.56	1.1	0.33
Bwe Karen	**20**	**0**	NA	8.6	4.4	5.3	1.8	**24.3**	**0.44**	0.33	0.93	0.61	0.24	0.55	0
White Hmong	**28.8**	**0**	41.36	12.7	4.6	8.5	3.1	**22.3**	**0.72**	0.53	1.52	1.04	1.0	0.58	0
Sedang	**13.6**	**11.9**	NA	5.2	0.8	6.8	0.8	**26.9**	**1.03**	1.24	1.13	2.16	0.45	0.83	0
Tetun	**34.5**	**7.7**	NA	16.1	1.9	13.5	3.1	**19.4**	**1.14**	1.29	1.57	1.76	1.35	0.95	0.16
Yay	**20.8**	**0**	NA	6.3	4.1	10.3	0	**18.1**	**1.10**	0.94	1.63	1.51	2.22	0.76	0
CORE															
Buli	**20.6**	**0**	NA	7.2	3.5	9.4	0.6	**44.3**	**1.05**	1.06	2.32	1.36	0.67	0.96	0.5
Efik	**47.6**	**32.8**	NA	23.8	9.6	10.4	3.8	**25.5**	**0.55**	0.31	3.64	0.69	0.15	0.87	0.2
Kanuri	**20.9**	**17.0**	37.94	7.5	3.5	7.5	2.4	**2.0**	**1.29**	0.63	4.0	15.0	0.8	0.73	0
Dongolese Nubian	**7.5**	**0**	NA	2.1	0.8	4.6	0	**34.3**	**0.59**	1.86	1.27	0.75	0.33	0.56	0.5
Rendille	**4.9**	**0**	NA	2.2	0	2.7	0	**30.2**	**0.44**	0.83	1.41	0.27	1.0	0.5	0.5
Burarra	**18.8**	**79.9**	NA	8.3	2.1	4.2	4.2	**40.6**	**1.34**	1.28	2.22	1.92	0.5	1.46	0.75
Kyaka	**22.1**	**0**	NA	7.5	3.7	7.2	3.7	**32.3**	**1.10**	0.57	3.78	1.8	0.29	0.84	1.0
Nunggubuyu	**1.4**	**66.7**	NA	1.4	0	0	0	**34.8**	**2.12**	2.0	3.82	2.29	NA	1.89	3.0
Kosarek Yale	**14.1**	**6.5**	NA	4.8	0.9	7.4	0.9	**35.1**	**1.29**	1.17	2.35	1.87	0.5	1.08	0.5
Greek	**5.2**	**36**	NA	2.5	0.6	1.1	0.9	**22.7**	**1.32**	2.33	1.72	1.66	3.5	0.85	0.5
Sora	**29.8**	**24.9**	NA	10.3	5.0	12.1	2.5	**25.2**	**0.79**	0.4	2.1	0.7	0.6	1.2	0.8
Wintu	**14.4**	**21.9**	NA	5.6	4.3	4.1	0.5	**39.7**	**0.70**	0.39	1.89	1.25	0.5	0.93	0
Hawaiian	**33.7**	**21.9**	29.22*	14.3	5.7	10.0	3.8	**37.3**	**0.98**	0.5	4.59	1.47	0.7	0.74	0.64

App. B (cont'd)

Manange	**14.9**	**5.6**	27.6	6.6	0.8	5.0	2.5	**21.5**	**0.59**	0.31	1.6	0.86	3.0	0.4	0
Mandarin	**17.2**	**0**	71.02	6.0	3.0	7.0	1.1	**20.2**	**0.96**	0.6	1.73	0.86	3.0	0.87	0
Vietnamese	**19.1**	**0**	38.38	8.4	3.8	5.3	1.5	**6.1**	**0.93**	0.58	1.0	1.5	0.33	1.0	0
EXT-1															
Anggor	**15.3**	**0**	NA	4.7	4.7	5.9	0	**21.6**	**1.68**	2.0	2.33	3.8	2.25	0.8	0
Dadibi	**27.6**	**8.5**	NA	10	5.9	9.4	2.4	**14.1**	**0.36**	0.43	1.18	0.95	0	0.14	0
Kemtuik	**0**	**0**	NA	0	0	0	0	**0**	**NA**	NA	NA	NA	NA	NA	NA
Lavukaleve	**3.2**	**33.3**	NA	1.1	0	2.1	0	**36.7**	**1.58**	0.5	2.14	1.5	3.0	1.6	NA
One	**25**	**0**	NA	9.0	2.1	12.8	1.1	**16.0**	**1.48**	1.06	2	1.6	NA	0.62	NA
Rao	**2.5**	**0**	NA	0	2.5	0	0	**5.0**	**0**	0	NA	NA	0	NA	NA
Sko	**23.5**	**8.7**	NA	9.2	4.1	9.2	1	**16.3**	**0.8**	0.67	0.88	0.53	3.5	1.1	0
Sentani	**7.1**	**0**	NA	3.2	0	1.6	2.4	**24.1**	**1.03**	0	2.5	1.8	NA	0.33	0
Tasmanian	**0**	**0**	NA	0	0	0	0	**39.7**	**0.87**	NA	1.69	1.6	NA	0.47	2
Waris	**8.1**	**0**	NA	4.1	0	4.1	0	**24.1**	**1**	0.2	1.97	1.75	0	0.6	NA
Yei	**8.1**	**0**	NA	5.4	0	2.7	0	**0**	**2**	0	NA	NA	NA	0	NA
Acoma	**6.3**	**100**	NA	1.6	3.2	1.6	0	**11.1**	**0.38**	1.0	0.27	0	0.5	1.0	NA
Quileute	**3.3**	**12.5**	NA	1.7	1.3	0	0.4	**20.1**	**0.6**	0.33	1.2	0.87	0.6	0.5	0
Abipón	**31.6**	**85.5**	NA	13.1	12.9	4.6	1	**23.0**	**0.21**	0.1	1.41	0.44	0.09	0.06	0
Lengua	**8.4**	**6.3**	NA	1.1	4.7	1.6	1.1	**31.4**	**0.67**	0	1.93	1.05	1.14	0.1	0
Macaguán	**3.7**	**66.7**	NA	1.2	0	2.5	0	**21.0**	**0.52**	0	0.79	0.69	0	0.6	0
Sáliba	**3**	**0**	NA	3	0	0	0	**11.0**	**1.83**	0.5	4.5	2.75	0	NA	NA
Tehuelche	**14.7**	**64.1**	NA	2.9	4.8	5.1	1.9	**35.3**	**0.27**	0.08	0.95	0.54	0	0.12	0
EXT-2															
Bakueri	**5.9**	**14.3**	NA	3	0.4	2.5	0	**11.0**	**2.22**	1	5.5	3.25	0	1.5	0
Koyraboro Senni	**13.6**	**8.2**	NA	2.5	3.0	8.1	0	**26.7**	**0.65**	0.69	1.82	0.81	1	0.45	NA

App. B (cont'd)

Noni	**16.5**	**30.2**	NA	8.1	3.3	2.3	2.8	**20.6**	**1.05**	0.49	1.75	0.91	2.0	0.62	NA
Swahili	**20.2**	**3.8**	29.7	9.2	3.4	5.7	1.9	**17.7**	**0.45**	0.18	1.27	0.91	0.15	0.38	0
Yoruba	**27.5**	**5.2**	NA	13.4	5.9	6.3	2.0	**21.7**	**0.78**	0.56	1.48	0.7	0.52	1.55	0.17
Gurindji	**5.8**	**71.4**	42.16	0.8	3.3	1.7	0	**40.0**	**0.43**	0.25	1.13	0.56	0.75	0.38	0
Muna	**15.6**	**40.5**	NA	5.6	3.0	5.1	2.0	**31.8**	**0.88**	0.76	1.63	1.62	0.51	0.8	0.22
Ngaanyatjarra	**4.7**	**8.1**	NA	0.8	0.9	0.8	2.3	**47.1**	**0.57**	0.72	1.94	0.64	0.71	0.82	0
Japanese	**25.9**	**12.5**	49.9	9.5	7.7	6.5	2.2	**8.5**	**0.64**	0.38	1.09	0.55	0.33	1.37	0
Blackfoot	**22.1**	**26.4**	NA	7.5	8.8	4.2	1.7	**10.8**	**0.56**	0.34	1.89	1.27	0.56	1.0	0
Cahuilla	**14.0**	**63.5**	NA	5.5	1.7	5.1	1.7	**22.9**	**0.30**	0.18	0.83	0.17	0.17	0.55	0.25
Comanche	**18.0**	**12.2**	NA	5.3	9.6	2.2	0.9	**20.0**	**0.39**	0.54	0.76	NA	0.33	0.22	2.0
Kashaya	**32.6**	**6.4**	NA	10.5	7.2	12.3	2.5	**11.0**	**0.52**	0.43	0.79	0.62	0.22	0.7	0.75
Lake Miwok	**15.4**	**27.8**	NA	4.3	3.4	5.6	2.1	**31.3**	**0.99**	0.48	1.97	1.13	0.6	1.55	0.25
Lakhota	**18.4**	**0**	NA	6.3	6.9	4.0	1.1	**16.1**	**0.84**	0.8	1.8	1.18	0.25	1.5	0
Lesser Antillean Creole French	**12**	**0**	NA	4.1	0	5.2	2.6	**33.8**	**1.61**	1.28	3.19	2.3	5.0	1.0	0.67
Tuscarora	**24.9**	**30.2**	NA	13	6.8	2.8	2.4	**29.7**	**0.66**	0.39	2.91	0.83	0.3	1.44	0.17
Yana	**16.4**	**12.5**	NA	8.2	2.7	4.1	1.4	**10.4**	**0.39**	0.5	0.07	0.64	0	0.62	0
Yaqui	**16.3**	**9.3**	36.56	8.8	1.9	4.8	0.8	**18.4**	**0.77**	0.36	2.04	0.78	0	1.27	0
Yuki	**19**	**25.2**	NA	6.8	7.4	4.8	0	**18.1**	**0.38**	0.28	0.98	1.54	0.27	0.2	0
San Lucas Quiaviní Zapotec	**8.6**	**4.5**	NA	3.1	0.8	4.7	0	**31.4**	**0.84**	0.4	1.54	1.02	1.5	0.71	0
Huambisa	**4.5**	**0**	NA	1.3	0.9	1.8	0.6	**30.1**	**0.50**	0	1.15	0.5	0.25	0.75	0
Kaingang	**25.9**	**5.2**	NA	9.8	5.4	8.9	1.8	**20.5**	**0.73**	1.27	0.48	0.91	0.44	0.73	0.4

App. B (cont'd)

Maxakalí	**37.3**	**0**	NA	16.5	7.0	10.1	3.8	**27.2**	**1.33**	0.7	6.17	4.29	0.5	0.67	0.67
Ancash Quechua	**11.1**	**28**	NA	2.3	2.5	4	2.3	**37.6**	**1.34**	1.04	2.76	3.63	0	0.95	0.64
Toba	**22.3**	**53.1**	NA	7.8	6.2	7.6	0.8	**18.4**	**0.77**	0.4	2.73	0.77	0.3	1.88	0
Tsafiki	**32.5**	**1.5**	NA	9.2	2.9	18.0	2.4	**21.8**	**1.96**	1.88	3.5	4.83	0.5	1.46	0.67

App. B (cont'd)

Wayampi	**26.9**	**14.9**	NA	14.4	2.5	8.1	1.9	**28.5**	**0.63**	0.62	1.44	0.81	0.63	0.45	0.43
Fijian	**36.7**	**29.2**	NA	14.5	7.3	12	2.9	**28.1**	**1.57**	0.92	4.51	2.68	0.26	1.73	1.29
Hani	**19.7**	**28.2**	NA	8.4	1.3	6.9	3.1	**18.5**	**0.95**	0.62	1.54	1.69	0.54	0.71	0.29
Kapingamarangi	**31.7**	**24.2**	NA	13.5	2.9	13.5	1.9	**20.7**	**0.79**	0.47	2.91	0.84	0.67	0.75	1.0
Lenakel	**21.3**	**6.5**	NA	7.8	0.8	12.0	0.8	**29.1**	**1.92**	1.23	3.55	2.58	1.0	1.85	1.0
Malagasy	**30.5**	**12.5**	53.74	10.3	9.2	8.8	2.3	**20.6**	**1**	0.66	2.86	1.69	0.1	1.53	0
Rotuman	**15.3**	**13.4**	NA	5.0	4.8	4.8	0.7	**37.1**	**0.67**	0.32	2.78	0.81	0.86	0.36	2.0
Samoan	**35.2**	**26.9**	NA	12.4	7.2	12.3	3.3	**27.9**	**0.91**	0.43	4.0	1.35	0.43	1.42	0.23
Takia	**18.7**	**0**	25.02	8.8	0.8	7.6	1.5	**20.9**	**1.29**	0.95	3.32	0.96	0	2.5	NA

* based on different sources

Appendix C: Structural and Sociocultural Data

Language	Consonant Inventory †	Syllable Structure ††	Vowel Inventory †††	Tone ‡	Root Structure ‡‡	Consonant-Vowel Ratio ‡‡‡	CPCI	Verbal Person Marking ✝	Core Cases ✝✝	Second Language Learners	Speakers	Percentage of Loanwords with European origin in Artifacts	European influence ✠	Hammarström ✠ ✠	Murdock/White & Levinson ✠ ✠ ✠	Noun/Verb ʘ
STATISTICS																
Hausa	4	2	2	2	3	4	44.33	0	1	1	24,162,000	NC	NC	A	Ad	0/0
Katcha	3	NA	3	3	4	3	48	NA	0	0	81,500	NC	NC	A	NA	0/0
Khoekhoe	4	2	2	3	2	4	34.33	NA	1	0	233,701	NC	NC	A	D	50/50
Mbum	4	2	2	2	2	4	39.33	0	NA	0	51,100	NC	NC	A	NA	0/0
Ngambay	3	2	3	3	2	2	35.33	1	0	1	750,000	NC	NC	A	NA	0/11.1
Baruya	1	2	3	NA	4	1	37.33	NA	1	0	6,600	NC	NC	A	Ghf	0/0
Berik	NA	NA	NA	NA	2	NA	NA	NA	1	NA	1,200	NC	NC	H	NA	0/0
Buin	1	3	2	1	4	1	44	2	1	0	26,500	NC	NC	A	NA	10/10
Kaluli	1	2	3	NA	2	1	27.33	NA	1	0	2,500	NC	NC	H	Bh	0/0
Kwoma	3	2	3	NA	3	3	40.33	NA	1	1	3,000	NC	NC	H	Bg	0/0
Mali	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	2,200	NC	NC	A	NA	0/0
Meyah	1	3	2	3	3	3	39	1	0	0	14,783	NC	NC	A	NA	0/0
Rotokas	1	1	2	1	NA	1	16	NA	NA	0	4,320	NC	NC	A	NA	37.5/50
Sahu	3	2	2	1	4	3	45.33	2	0	0	7,500	NC	NC	A	NA	0/22.2
Toaripi	1	1	2	1	NA	1	16	0	1	0	23,000	NC	NC	H	BGH	0/0

App. C (cont'd)

Yir Yoront	3	NA	2	1	1	3	25.5	0	1	1	45	NC	NC	H	NA	0/0
Abzakh Adyghe	5	3	1	NA	1	5	45	2	1	0	499,180	NC	NC	A	Ae	0/0
Badaga	3	3	2	1	3	3	47	1	1	0	245,378	NC	NC	A	Bh	0/10
Basque	3	3	2	1	3	4	47	2	1	0	588,108	NC	NC	A	Ad	0/0
Bezhta	5	3	2	1	2	5	50	0	1	0	3,000	NC	NC	A	NA	0/0
Chukchi	2	3	2	1	3	2	43	2	1	1	10,000	NC	NC	H	Df	11.1/11.1
Ket	2	3	3	1	2	2	38	2	0	0	770	NC	NC	H	HF	0/0
Khalkha	3	2	3	1	3	2	40.33	0	1	1	2,337,095	NC	NC	A	D	22.2/22.2
Laz	4	3	2	NA	3	4	51	2	1	0	33,000	NC	NC	A	ADe	12.5/12.5
Nivkh	4	3	2	2	1	4	41	0	0	0	1,089	NC	NC	H	Fh	0/11.1
Kildin Saami	4	3	2	1	NA	4	54	0	1	0	800	NC	NC	A	D	0/0
Welsh	3	3	3	1	2	2	42	1	0	0	536,258	NC	NC	A	Ad	0/0
Yukaghir	3	2	2	1	3	3	40.33	1	1	0	30	NC	NC	H	Hf	0/0
Biloxi	1	3	2	NA	3	1	39	2	0	NA	NA	0	iS/E/F	H	NA	0/40
Carrier	5	NA	2	2	1	NA	37.5	2	0	0	1,500	5.32	E/F	H	Hfe	0/0
Upper Chehalis	4	3	1	1	2	5	46	2	0	NA	NA	18.83	(E/F)	H	NA	55.5/55.5
Cheyenne	1	NA	1	2	NA	3	NA	2	0	0	1,720	0	E	H	Hgb	0/0
Chickasaw	2	2	1	1	3	4	36.33	2	1	0	1,000	3.85	iS/E/F	A	HFGB	10/30
Highland Chontal	4	3	2	1	4	4	56	2	0	0	3,600	73.33	(dS)	A	A	0/0
Ineseño Chumash	5	3	2	NA	2	4	50	2	0	NA	NA	47.37	(dS/E)	H	Gh	0/25
Haida	5	3	1	1	2	5	50	0	0	0	40	4.35	E/F	H	F	0/10
San Mateo del Mar Huave	3	2	2	2	1	3	30.33	2	0	0	12,000	55.26	dS	A	HDAFe	0/0
Itzaj	3	2	2	1	1	3	30.33	NA	0	0	12	63.48	(dS)	A	C	0/44.4
Kiliwa	3	2	1	3	1	4	30.33	2	0	0	28	0	iS	H	NA	0/0

App. C (cont'd)

Kiowa	3	2	2	2	1	3	30.33	NA	0	0	1,092	4.76	E	A	Hd	12.5/25
Nez Perce	4	3	2	1	3	4	51	2	1	0	200	0	E/F	H	NA	44.4/44.4
Nuuchahnulth	5	3	1	1	3	5	55	1	0	0	200	18.18	(E/F)	H	NA	42.86/57.14
Oneida	1	3	1	2	2	3	34	2	0	0	250	0	E/F/Du	A	Cfg	0/0
Santiago Mexquititlan Otomí	NA	NA	NA	NA	2	NA	NA	NA	NA	0	33,000	20.19	dS	A		0/0
Pawnee	1	3	1	1	2	1	34	2	0	0	20	5	(E)	A	Ch	0/11.1
Pipil	1	NA	2	1	2	3	21	2	0	0	20	81.25	dS	A	NA	0/14.26
Xicotepec de Juárez Totonac	NA	NA	NA	1	3	NA	NA	2	0	0	3,000	46.67	dS	A	A	0/0
Wappo	4	2	2	1	2	4	39.33	0	1	NA	NA	*59.52*	dS/E	H	NA	0/0
Central Yup'ik	3	2	1	1	3	4	40.33	2	1	0	1,000	14.8	R/E	A	F	0/10
Copainalá Zoque	2	3	2	1	2	2	38	2	1	0	10,000	9.09	dS	A	NA	0/0
Aguaruna	1	2	1	NA	3	3	32.33	2	1	1	38,290	38.88	?S	A	Chf	0/0
Arabela	4	2	2	1	NA	2	44	NA	0	0	50	8.33	(?S)	H	Cd	0/0
Aymara	4	2	1	1	3	5	44.33	2	1	0	22,227,642	78.57	dS	A	NA	0/0
Bora	3	2	2	2	4	3	45.33	1	1	0	2,828	5.54	(dS)	H	NA	0/25
Bororo	1	1	3	1	3	1	25.67	2	0	0	850	9.38	(?P)	H	HFC	0/0
Carib	2	2	2	1	4	2	41.33	2	0	0	10,226	*64*	dS/E/F	A	C	0/0
Cashinahua	1	2	1	2	2	3	27.33	NA	1	0	2,000	5.56	iS	A	HCgf	0/0
Cavineña	3	1	1	NA	4	4	38.67	0	1	0	1,180	50	?S	H	NA	0/20
Cayapa	3	2	1	NA	NA	4	38	NA	0	0	3,450	46.67	dS	A	H	0/0
Chayahuita	1	1	1	NA	NA	3	16	NA	1	0	11,384	54.5	?S	H	Behf	40/40
Cubeo	1	1	2	2	NA	1	16	1	1	1	6,150	5.55	?S	A	Cf	0/0
Embera	2	2	2	1	3	2	36.33	0	1	0	23,480	41.18	(?S)	A	Bhfg	0/0
Guaraní	3	1	2	1	3	3	33.67	2	0	1	4,848,000	9.32	dS	A	NA	0/12.5
Hupda	3	2	3	2	1	1	30.33	0	1	0	1,358	26	(iP)	H	NA	0/0

App. C (cont'd)

Jarawara	1	1	1	1	4	3	30.67	2	0	0	155	68.18	(dP)	H	NA	0/0
Miskito	1	2	1	NA	3	3	32.33	1	0	1	183,400	*33.54*	dS/E	A	NA	0/0
Piro	2	3	2	2	3	3	43	2	0	0	4,000	23.17	?S	A	H	0/0
Imbabura Quechua	3	2	1	1	3	5	40.33	2	1	1	300,000	61.54	(dS)	A	NA	0/0
Rama	2	2	1	1	2	4	31.33	1	1	0	24	0	(?E)	A	NA	0/0
Wichí	3	2	2	1	3	3	40.33	2	1	0	15,000	36	?S	H	NA	0/0
Yanomámi	1	1	3	1	4	1	30.67	NA	1	0	9,000	8.33	(?S/?P	H	C	0/0
Great Andamanese	1	2	3	1	2	3	27.33	NA	1	NA	NA	NC	NC	H	Fg	0/12.5
Bislama	2	3	2	1	2	3	38	0	0	1	6,200	NC	NC	A	NA	0/30
Bwe Karen	4	2	3	3	1	3	34.33	NA	NA	NA	15,700	NC	NC	A	NA	37.5/50
White Hmong	5	2	3	3	1	5	38.33	0	0	0	818,685	NC	NC	A	HA	0/14.26
Sedang	5	2	3	1	1	4	38.33	0	0	0	101,434	NC	NC	A	H	0/0
Tetun	1	2	2	1	3	3	32.33	0	0	1	450,000	NC	NC	A	Cd	10/30
Yay	3	2	2	3	1	3	30.33	0	0	0	2,049203	NC	NC	A	NA	0/0
CORE																
Buli	3	2	3	NC	2	NC	35.33	NC	0	NC	NC	NC	NC	A	NA	0/12.5
Efik	1	2	3	NA	NA	NC	26	NC	0	NC	NC	NC	NC	A	NA	0/0
Kanuri	4	2	3	NC	NC	3	NC	NC	1	NC	NC	NC	NC	A	AE	0/12.5
Dongolese Nubian	3	2	3	1	2	NC	35.33	NC	1	NC	NC	NC	NC	A	AD	14.26/14.26
Rendille	3	2	2	NC	2	NC	42	NC	1	NC	NC	NC	NC	A	NA	0/0
Burarra	2	3	2	NC	NC	3	NC	NC	0	NC	NC	NC	NC	H	NA	0/0
Kyaka	NC	NC	NC	NC	NC	NC	NC	NC	1	NC	NC	NC	NC	A	NA	0/0
Nunggubuyu	NC	NC	NC	NC	NC	NC	NC	NC	1	NC	NC	NC	NC	H	NA	0/0
Kosarek Yale	1	NA	2	NC	NC	NA	NC	NC	0	NC	NC	NC	NC	A	NA	0/0
Greek	3	3	2	1	3	NC	47	NC	1	NC	NC	NC	NC	A	Adt	0/0

App. C (cont'd)

Sora	NC	NC	NC	NC	NC	NC	NC	NC	1	NC	NC	NC	NC	A	Bc	37.5/37.5
Wintu	4	2	2	NC	NC	4	NC	NC	1	NC	NC	22.42	iS/E	H	NA	0/20
Hawaiian	1	1	2	1	3	NC	25.67	NC	1	NC	NC	NC	NC	A	NA	0/60
Manange	4	3	2	3	1	NC	41	NC	NA	NC	NC	NC	NC	A	NA	0/0
Mandarin	4	3	2	3	1	NC	34.33	NC	0	NC	NC	NC	NC	A	NA	0/0
Vietnamese	3	2	3	3	1	NA	30.33	NC	0	NC	NC	NC	NC	A	A	0/0

Ext-1

Anggor	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NA	NA	0/14.26
Dadibi	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	Chg	0/0
Kemtuik	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Lavukaleve	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/0
One	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/0
Rao	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Sentani	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Sko	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Tasmanian	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	H	0/0
Waris	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/0
Yei	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Acoma	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	Bd	100/100
Quileute	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/12.5
Abipón	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	Hd	12.5/12.5
Lengua	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	H	0/0
Macaguán	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Sáliba	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	Chf	0/0
Tehuelche	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/11.1
Toba	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	10.87	?S	H	HA	11.1/11.1

App. C (cont'd)

EXT-2

Bakueri	3	NA	3	3	2	NC	33	NC	NC	NC	NC	NC	NC	A	NA	0/0
Koyraboro Senni	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/33.3
Noni	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/0
Swahili	4	1	2	1	3	4	37.67	NC	NC	NC	NC	NC	NC	A	E	0/0
Yoruba	2	1	3	3	4	NA	34.67	NC	NC	NC	NC	NC	NC	A	Cefh	12.5/12.5
Gurindji	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/0
Muna	3	1	2	1	3	NC	33.67	NC	NC	NC	NC	NC	NC	A	NA	0/11.1
Ngaanyatjarra	2	NA	2	1	4	NA	42	NC	NC	NC	NC	NC	NC	H	H	0/0
Japanese	2	2	2	NC	NC	2	NC	NC	NC	NC	NC	NC	NC	A	L	28.6/28.6
Blackfoot	1	NA	1	NC	NC	3	NC	NC	NC	NC	NC	NC	NC	H	Hg	0/0
Cahuilla	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	H	42.86/42.86
Comanche	1	2	2	1	4	NC	37.33	NC	NC	NC	NC	4.55	iS/E/F	A	Hd	0/0
Kashaya	4	2	2	NC	NC	4	NC	NC	NC	NC	NC	*28.25*	(R/?S/E)	H	Gh	0/0
Lake Miwok	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	59.82	?S/E	H	NA	0/50
Lakhota	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	Hf	0/0
Lesser Antillean Creole French	3	3	3	NC	NC	3	NC	NC	NC	NC	NC	NC	NC	A	A	0/0
Tuscarora	1	3	2	1	1	NC	29	NC	NC	NC	NC	*0*	(E/F)	A	NA	0/10
Yana	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	25/25
Yaqui	2	2	2	2	3	NC	36.33	NC	NC	NC	NC	67.3	(dS)	A	A	0/0
Yuki	3	3	2	NC	NC	3	NC	NC	NC	NC	NC	10.87	iS/E	H	NA	0/12.5
San Lucas Quiaviní Zapotec	4	3	2	NC	NC	3	NC	NC	NC	NC	NC	72.92	(dS)	A	Ad	11.1/11.1
Huambisa	1	NA	2	NA	3	NC	28.5	NC	NC	NC	NC	*58.33*	?S	A	C	0/0
Kaingang	1	3	3	2	2	NC	34	NC	NC	NC	NC	20.59	P	NA	BHg	0/0
Maxakalí	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	H	NA	0/50

App. C (cont'd)

Ancash Quechua	2	NA	2	NA	3	NC	34.5	NC	NC	NC	NC	*50*	dS	A	NA	10/20
Tsafiki	2	1	2	1	2	NC	24.67	NC	NC	NC	NC	*46.15*	dS	A	Befh	0/0
Wayampi	1	1	2	1	3	NC	25.67	NC	NC	NC	NC	*20*	(?S/?F)	A	Chf	0/22.2
Fijian	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/70
Hani	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	0/0
Kapingamarangi	1	NA	2	1	3	NC	28.5	NC	NC	NC	NC	NC	NC	A	NA	0/57.14
Lenakel	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	A	NA	20/20
Malagasy	4	1	2	1	4	NC	42.67	NC	NC	NC	NC	NC	NC	A	NA	0/0
Rotuman	1	2	2	1	3	NC	32.33	NC	NC	NC	NC	NC	NC	A	Bf	0/30
Samoan	1	1	2	1	3	NC	25.67	NC	NC	NC	NC	NC	NC	A	Bd	0/30
Takia	2	3	2	1	2	NC	38	NC	NC	NC	NC	NC	NC	A	NA	0/0

Legend:

General: NA: no data available; NC: no data coded

†	1: Small, 2: Moderately small, 3: Average, 4: Moderately large, 5: Large
††	1: Simple, 2: Moderately complex, 3: complex
†††	1: Small, 2: Average, 3: Large
‡	1: No tones, 2: Simple tone system, 3: Complex tone system
‡‡	1: Short, 2: Moderately short, 3: Moderately long, 4: Long
‡‡‡	1: Low, 2: Moderately low, 3: Average, 4: Moderately high, 5: High
✝	0: No verbal person marking, 1: Only A, 2: Both A and P
✝✝	0: No core cases present, 1: core cases present
✠	d: direct influence, i: indirect influence, Du: Dutch, E: English, F: French, P: Portuguese, R: Russian, S: Spanish. Brackets indicate that data were added for the present study. When brackets are absent, data are from Brown (1999), who uses the same coding scheme.
✠ ✠	A: Agriculturalists, H: Hunter-Gatherers
✠ ✠ ✠	A: Advanced Agriculture, B: Horticulture, C: Simple or Shifting Cultivation, D: Domestic Animals, E: Exchange Economy, F: Fishing, G: Gathering, H: Hunting. Capital letters are used to indicate the dominant mode of subsistence, lower-case letters to indicate subsidiary mode(s) (not used for statistical analysis).
ʘ	The figure before the slash gives percentages only for derived terms, while that behind the slash also takes into account terms that are ambiguous between a nominal and a verbal reading.

Appendix D: Meanings and their Formal and Semantic Properties

Meaning	**Sum**	**Percentage Overt Marking in this work**	**Percentage Colexification**	MSC	SC	FC	TC	PC	PRC	CC	CS-Ratio	PS	FS	PF-Ratio	Common and areal associations
TOPOLOGICAL AND NATURE-RELATED TERMS															
1. 'animal'	**43.1**	**26.8**	18.3	20	0.4	4.1	0	11.1	5.3	3	0.1	1.6	0	0	C: 'thing,' 'meat', 'life/be alive' A: 'pig-dog' (New Guinea), 'move on earth' (Eastern North America)
2. 'ashes'	**38.5**	**17.4**	21.0	16	1.6	0	0.4	2.3	0.9	0	3.6	18.8	0	0	C: 'dust,' 'fire'
3. 'bark'	**68.0**	**23.7**	44.6	26	0.2	0.4	0	0	0	0	111	0.3	66.8	276.5	C: 'skin,' 'peel/rind/shell,' 'tree/wood,' 'hide/leather,' 'cover,' 'scale'
4. 'bay'	**90.3**	**47.3**	44.8	24	4.3	6.4	0	9.6	0	3.2	0.9	19.3	1.1	0.1	C: 'corner,' 'sea,' 'harbor,' 'lake' A: 'breast' (Europe)
5. 'beak'	**66.8**	**15.9**	52.4	28	0	2.8	0	0.5	0	0	17.8	38.1	19.1	0.5	C: 'nose,' 'mouth,' 'snout/muzzle,' 'end/point,' 'bird,' 'tooth' A: 'peck' (Western North America), 'prow of canoe' (South America)
6. 'bird'	**20.9**	**10.9**	10.7	8	0	0	0	9.2	0	0	0.2	1.6	0	0	C: 'animal,' 'chicken/fowl'
7. 'blossom'	**23.4**	**7.7**	15.7	8	0	1.1	0	2.9	0.2	0	1.8	7.3	0	0	A: 'picture' (Eurasia)

App. D (cont'd)

8. 'branch'	**40.0**	**12.5**	27.5	8	6.9	0	0	2.6	0.4	0	2.9	28.3	0	0	C: 'arm,' 'tree,' 'hand,' 'leg' A: 'knot in tree/knot in wood' (Western North America)
9. 'bud'	**53.5**	**19.7**	33.8	NA	0.4	0	1.3	5.5	0	3.3	0.8	7.9	0	0	C: 'shoot/sprout,' 'flower/blossom,' ' to sprout'
10. 'cave'	**53.3**	**24.3**	29.2	26	3.2	4.0	0	30.3	0.4	0	0.1	2.1	2.4	1.1	C: 'hole,' 'stone/rock,' 'den/lair'
11. 'clearing'	**56.0**	**42.7**	13.3	NA	6.4	0	0	37.7	1.6	0	0.1	2.7	0	0	C: 'open/clear/unobstructed,' 'valley,' 'cut' A: 'field/meadow/ lawn' (North America)
12. 'cloud'	**41.2**	**14.6**	27.4	12	7	0	0	1.8	1.8	0	0.6	26.6	0	0	C: 'fog/mist,' 'sky,' 'smoke' A: 'smoke' (South America)
13. 'coal'	**29.3**	**12.8**	16.4	24	0.2	0.3	0.2	3.6	1.4	0.2	0.6	3.6	0	0	C: 'embers,' 'fire,' 'black'
14. 'coast'	**53.0**	**41.2**	12.4	48‡	11.7	0	0	3.8	0	1.7	1.3	22.7	0	0	C: 'edge/end/border/ point/limit,' 'sea,' 'water'
15. 'dew'	**30.0**	**16.3**	14.1	20	0.4	0	0	1.8	0.9	2.9	1.4	13.5	7.9	0	C: 'water,' 'fog,' 'cold/frost' A: 'wet/moist' (Western North America)
16. 'dust'	**54.2**	**12.4**	41.8	14	0	2.7	2	0	0.9	0	3.5	19.3	0	0	C: 'powder/grit,' 'land/soil,' 'earth,' 'dirt/rubbish/garbage,' 'ashes,' sand' A: 'smoke' (South America)
17. 'eclipse'	**72.2**	**63.1**	9.2	NA	0	0	0	12.8	0	0	4.5	56.9	0	0	C: 'sun,' 'moon,' die/kill,' 'darkness/dark,' 'eat'
18. 'egg'	**26.2**	**5.2**	21.0	10	0	0.7	0	0	0.7	0	0.9	1.3	0	0	C: 'testicles' A: 'nut' (New Guinea)
19. 'embers'	**55.9**	**26.2**	30.2	34	0.4	0.5	0	8	0	7.6	0.4	6.4	0.5	0.1	C: 'coal,' 'fire,' 'flame,' 'ashes,' 'burn'

App. D (cont'd)

20. 'estuary'	**49.6**	**36.5**	15.2	100 *	4.6	0	0	5.3	0	0	2.7	27.0	0	0	C: 'mouth,' 'river/stream,' opening,' 'water' A: 'sea' (Europe)
21. 'feather'	**56.0**	**11.2**	45.0	24	9.3	1.1	0	0	0	0	3.9	1.7	38.9	23.4	C: 'hair,' 'fur/wool,' 'wing,' 'leaf,' 'bird' A: 'beard' (Mainland Southeast Asia and Oceania), 'pen' (Old World)
22. 'flame'	**63.2**	**36.3**	27.8	48	0	0	0	18.5	0	8.6	1.0	26.2	0	0	C: 'fire,' 'tongue,' 'light', 'burn'
23. 'flood'	**49.6**	**30.6**	18.5	NA	0	0	0	9	0	19	0.4	10.2	0	0	C: 'water,' 'river/stream,' 'torrent'
24. 'foam'	**32.5**	**11.5**	21.0	10	0	0	0	11.5	0.9	1.2	1.0	13.9	0	0	C: 'bubbles' A: 'saliva' (Mainland Southeast Asia and Oceania)
25. 'fog/mist'	**47.7**	**13.4**	34.5	18	0.4	0	0	1.5	0	0	18.3	35.3	0	0	C: 'cloud,' 'smoke,' 'steam,' 'dew' A: 'darkness' (Europe and Eurasia)
26. 'forest'	**42.1**	**10.9**	31.3	12	4.9	0	0	0.2	7.1	9.1	0	0	0	NA	C: 'bush/grove/scrub,' 'wood, 'tree,' 'mountain/hill,' A: 'taiga' (Eurasia)
27. 'gold'	**34.2**	**20.3**	14.0	0	6.5	5.8	0	11.3	0	0	0.4	6.8	1.1	0.2	C: 'yellow,' 'metal,' 'money'
28. 'grass'	**37.2**	**2.9**	34.2	6	0.8	0.4	0	0	11.6	6.3	0	0.8	0	0	C: 'hay/straw,' 'meadow/pasture/lawn,' 'plant'
29. 'headland'	**47.0**	**30.8**	18.8	NA	0	0	0	10.3	0	0	2.9	29.5	0	0	C: 'nose,' 'point/tip,' 'head/ forehead,' 'end/to end'
30. 'honey'	**53.2**	**27.7**	25.4	18	1.1	0	0	8.5	22.2	0	0.4	14.1	0	0	C: 'bee,' 'liquid/sap/juice,' 'sweets/sugar'
31. 'horizon'	**66.5**	**67.0**	1.9	NA	3.7	0	0	0	0	0	15	54.9	0	0	C: 'sky,' 'edge/border/fringe,' 'end/finish,' 'land/earth,' 'cloud,' 'meet/meeting place,' 'basis'
32. 'horn'	**44.7**	**7.2**	37.5	2	0	0	0	0	0	12.2	2	23.7	0	0	C: 'antlers'
33. 'lagoon'	**51.5**	**23.5**	28.9	24	1.5	0	1.8	6.6	0	6	1.9	29.5	0	0	C: 'lake/pond,' 'water/liquid,' 'sea'

App. D (cont'd)

34. 'lake'	**45.0**	**14.6**	31.7	16	0.1	0	0	1.5	16.1	1.7	0.6	0	0	0	C: 'puddle,' 'lagoon,' 'water,' 'river,' 'swamp,' 'sea'
35. 'lightning'	**39.9**	**22.0**	18.4	43†	0	0	6.9	4.6	0.7	0	1.8	22	0	0	C: 'thunder,' 'gleam/lighten /shine' A: 'arrow' (Eurasia)
36. 'meteoroid (shooting/shining star)'	**61.5**	**55.9**	5.6	NA	0	0	0	2.1	0	3.2	10.3	54.3	0	0	
37. 'Milky Way'	**70.3**	**65.3**	5.9	NA	0	0	0	1.8	0	0	37.2	64.7	1.8	0	C: 'trail/road/street,' 'star,' 'sky,' 'river'
38. 'moon'	**67.1**	**6.5**	60.6	8	0	0.7	48.5	2.6	0.7	0	0.1	5.2	0	0	C: 'month,' 'sun,' A: 'snail' (Australia)
39. 'mountain'	**53.7**	**11.4**	42.7	6	11.1	0	0	0	0	4.9	1.8	28.6	0	0	C: 'hill'
40. 'mushroom (fungus)'	**17.8**	**15.4**	2.9	24	0	0	0	1	0	0	16.0	15.7	0	0	
41. 'nest'	**40.0**	**21.0**	19.0	20	0	1.8	0	1	2.3	0	6.3	0.6	31.5	49.5	C: 'house/home,' 'bird,' 'den/lair' A: 'raft' (Australia)
42. 'plant'	**67.4**	**20.6**	46.8	50	4.4	6.7	0	12.6	3.2	1.7	0	0	0	NA	C: 'tree,' 'grass/weed,' 'thing,' 'grow,' 'vegetable'
43. 'puddle'	**85.6**	**68.1**	19.6	NA	1.5	0.4	1.5	6.6	2.9	19.5	0.1	3.4	0	0	C: 'pond/lake,' 'water,' 'swamp,' 'stand/sit/be stagnant,' 'hole,' 'mud'
44. 'rain'	**28.9**	**8.8**	21.4	10	0	0	5.1	1.4	5.8	9.1	0.1	2.7	0	0	C: 'water/liquid'
45. 'rainbow'	**33.7**	**28.8**	6.3	42	0.8	0	0.4	2.3	0	0	9.1	30.8	0	0	C: 'snake,' 'bow/arc/bend' A: 'thunder' (Eurasia)
46. 'resin'	**49.5**	**27.0**	24.6	NA	0	1	0	2.9	1	0	7.1	34.2	1	0	C: 'water/liquid/juice,' 'tree' A: 'birdlime' (Mesoamerica)

App. D (cont'd)

47. 'river/stream'	**43.6**	**11.0**	33.2	100 *	3.6	0.4	0	0	1.6	23.6	0.1	2.6	0	0	C: 'water,' 'spring/well,' 'lake/pond'
48. 'river bed'	**61.5**	**40.2**	25.2	NA	28.2	0	0	5.1	0	2.6	0.2	5.1	2.6	0.5	C: 'river/stream,' 'valley,' 'water,' 'way,' 'flow' A: 'deep' (Eurasia)
49. 'root'	**32.6**	**6.9**	25.8	6	3.8	0.6	0	0.9	0	0	1.6	5.7	2.7	0.5	
50. 'seed'	**43.3**	**6.4**	36.9	16	0	2.8	0	0.2	2.5	0	2.3	13.1	0	0	C: 'fruit' A: 'kind' (Africa)
51. 'shadow'	**49.3**	**7.7**	42.4	14	0	2.4	0	4.7	0	0	5.3	37.8	0	0	C: 'soul/spirit/ghost,' 'reflection/image/mirror,' 'image/picture/drawing,' 'photograph'
52. 'sky'	**49.1**	**11.8**	37.9	18	6.8	0	0	9	1.4	0	1.3	22.7	0	0	C: 'heaven,' 'high/above/up,' 'cloud'
53. 'smoke'	**33.9**	**7.0**	26.9	14	0	0	0	0.9	5.5	0.4	3.5	23.6	0	0	C: 'steam,' 'fog'
54. 'soil'	**72.4**	**6.4**	66.2	10	46.2	0	0	0.4	0.4	0	0	0.4	0	0	C: 'land/ground,' 'dirt,' 'world,' 'place'
55. 'spark'	**39.1**	**27.2**	13.1	NA	0	1.3	0	6	0.3	2.5	2.3	23.1	0	0	C: 'fire,' 'lightning'
56. 'spring/well'	**46.3**	**35.5**	12.0	30	0.8	1.9	0.3	2.6	2.5	19.8	0.4	12.5	0	0	C: 'water,' 'river,' 'hole,' 'eye'
57. 'star'	**21.2**	**5.2**	18.1	6	0	0.3	0	0.9	0	1.4	7.2	19.2	0	0	A: 'starfish' (Oceania)
58. 'steam'	**53.9**	**11.3**	42.7	24	0	0	0	1	15.5	1	1.7	29.4	0	0	C: 'smoke,' 'fog/mist,' 'breath/exhalation,' 'heat/hot'

App. D (cont'd)

59. 'straw'	**70.0**	**26.2**	43.8	0 *	0	6.2	0	0	34.1	13.1	0.1	6.9	0	0	C: 'grass,' 'dry' A: 'hat' (Eastern North America)
60. 'sun'	**51.8**	**5.8**	46.0	26	0	2.7	15.1	2	5.9	0	0.4	10.9	0	0	C: 'day,' 'moon,' 'clock'
61. 'swamp'	**42.0**	**25.3**	16.9	30	1.7	0	0	4.6	0.2	5.7	0.3	3.7	0	0	C: 'pool/pond/lake,' 'mud,' 'water,' 'puddle'
62. 'tail'	**21.5**	**3.8**	17.7	6	5.6	0	0	0	0	0	2.7	14.9	0	0	C: 'end'
63. 'thorn'	**42.2**	**7.8**	34.4	NA	0	2.8	0	3.3	1.9	0	2.3	18.4	0	0	C: 'needle/awl' A: 'sticker' (Western North America)
64. 'thunder'	**41.5**	**17.9**	24.5	36	0	0	12	0	0	0	2.3	28.1	0	0	C: 'lightning,' 'god/spirit'
65. 'tree'	**70.3**	**5.4**	65.3	10	0.8	0	0	0.7	48.5	0.8	0	0.9	0	0	C: 'wood,' 'stick,' 'plant,' 'trunk/log/pole'
66. 'valley'	**47.0**	**20.2**	27.8	22	17.5	0	0	5.3	1.2	0.9	0.3	7.9	0.3	0	C: 'gully/furrow/ditch/gorge/channel,' 'plain/low land,' 'river,' 'flat/flat land'
67. 'volcano'	**52.7**	**98.0**	2.0	NA	0.7	0	0	24.3	0	20	0.3	13.7	0	0	C: 'burn/fire,' 'mountain/hill,' 'earth/land'
68. 'waterfall'	**52.0**	**38.4**	13.6	64	1.9	0	0	4.3	0.5	24.7	0.2	6.6	0	0	C: 'current/rapids/cataract,' 'water,' 'fall' A: 'cliff' (South & Southeast Asia)
69. 'wave'	**19.8**	**17.5**	2.3	24	0	0	0	0.8	0.5	1.6	5.1	14.9	0	0	C: 'water'
70. 'wax'	**45.7**	**24.5**	21.2	62	0.6	1.2	0	0	13.4	0	1.7	23	2.5	0.1	C: 'candle,' 'bee,' 'honey'
71. 'whirlpool'	**50.1**	**43.1**	7.4	62	1.2	0	0	3.7	0	24.3	0.7	19.2	0	0	C: 'water,' 'go around/spin/turn/twist,' 'whirlwind'

App. D (cont'd)

Artifacts															
72. 'airplane'	**55.5**	**50.2**	5.2	44	1.8	1	0	1.9	0	0	6.1	10.4	18.4	1.8	C: 'fly,' 'boat/canoe,' 'bird, 'sky'
73. 'ball'	**31.7**	**12.5**	19.3	6	0	5.5	0	8.8	1.2	0	0.5	7.6	0	0	C: 'round'
74. 'bed'	**46.8**	**28.7**	17.7	18	28.2	0	0	5.1	0	2.6	0.3	5.1	2.6	0.5	C: 'sleep,' 'furniture,' 'place' A: 'lie/lie down' (North America)
75. 'belt'	**36.4**	**26.8**	9.3	30	7.9	12.7	0	1.5	0	0.8	0	0	0.3	NA	C: 'tie/bind/wrap/fasten,' 'waist,' 'fast/fasten'
76. 'boat'	**33.1**	**10.7**	23.3	10	1.3	3.8	0	1	0	1	1.3	3.4	5.4	1.6	C: 'canoe' A: 'collamon' (Australia)
77. 'car'	**41.9**	**27.8**	15.2	28	0	4.3	0	14.3	0	0	0.6	3.4	5.8	1.7	C: 'vehicle,' 'cart/carriage' A: 'move self,' 'machine' (Western North America)
78. 'chair'	**53.0**	**40.6**	12.8	26	2.3	34.6	0	0.8	0.8	0	0.3	10.3	0	0	C: 'sit down,' 'furniture,' 'wood/tree'
79. 'clock'	**60.0**	**27.6**	32.4	14	0	48.4	0	2.9	0	0	0.1	3.4	2.8	0.8	C: 'sun,' 'hour,' 'day,' 'time'
80. 'glasses'	**62.4**	**55.8**	6.6	44	16.1	11.6	0	0	0.6	2.7	0.8	23.4	1.2	0	C: 'eye,' 'glass,' 'mirror,' 'see/look'
81. 'house'	**29.7**	**3.0**	26.4	6	5	1.4	0	0.7	0.7	5.2	0.6	0.7	7	10.0	C: 'village' A: 'company/firm' (Europe), 'teepee' (Eastern North America), 'lineage' (Old World)
82. 'key'	**37.0**	**33.0**	5.3	34	0	33.5	0	0	0	0	0.1	3.3	0	0	C: 'open/close,' 'lock/keyhole,' 'lock/unlock'
83. 'knife'	**20.9**	**18.7**	2.5	12	0	8.1	0	0	5.7	0	0.9	0.5	12.3	24.4	A: 'machete' (South America)
84. 'ladder'	**52.9**	**33.3**	20.0	24	1.2	21.5	0	1.1	0.4	0	1.0	2.6	22.8	8.7	C: 'stairs/staircase,' 'climb/ascend,' 'step'
85. 'mirror'	**51.3**	**34.5**	16.8	30	0	20.5	0	2.4	1.3	0.3	0.2	4.8	0.4	0.1	C: 'glass/type of glass,' 'see,' 'look/inspect,' 'shadow' A: 'window' (North America)

App. D (cont'd)

86. 'needle'	**35.8**	**14.5**	21.6	12	0	11.4	0	0	0.8	0	0.9	3.8	7.2	1.9	C: 'sew,' 'injection/syringe'
87. 'paper'	**59.6**	**13.2**	46.8	0	0	7.6	0	0	2	38.4	0.1	4.4	0.8	0.2	C: 'letter/book/document'
88. 'pen'	**51.6**	**45.4**	6.2	20	0.2	38.5	0	0	1	0	0.1	1	2.4	2.3	C: 'write,' 'stick' A: 'poke' (Australia), 'feather' (Old World)
89. 'rope'	**55.5**	**5.1**	50.4	2	0	2.8	0	0	0	2.2	6.6	1.4	31	21.8	C: 'thread/string/cord/twine,' 'vine/climbing plant' A: 'lasso' (Mesoamerica)
90. 'scissors'	**28.6**	**24.8**	4.3	24	0	18.4	0	1	0	0	0.4	1	6.9	7.3	C: 'cut'
91. 'shoe'	**29.4**	**20**	9.4	10	15	4.8	0	0	1.1	0	0.1	0.8	0.3	0.3	C: 'foot'
92. 'street/road/way'	**26.2**	**9.4**	16.0	16	4	4.7	0	1.1	0	0	0.6	1.4	4.8	3.5	C: 'manner/method/system'
93. 'table'	**35.5**	**23.6**	11.9	8	0	18.6	0	0.9	0	0	0.4	7.9	0	0	C: 'eat,' 'furniture,' 'food' A: 'floor' (New Guinea)
94. 'toilet'	**54.4**	**49.3**	6.8	38	6.2	37.6	0	2.3	0	0	0.1	1.7	0.7	0.4	C: 'faeces,' 'house,' 'bathroom,' 'place'
95. 'train'	**52.2**	**45.6**	6.7	42	0	1.7	0	8.3	0	0	3.9	5.6	33.3	6.0	C: 'wagon/vehicle,' 'fire,' 'iron/metal/steel' A: 'run' (Eastern North America)
96. 'weapon'	**50.9**	**38.7**	14.2	42	0	32.7	0	2	0	0	0	0.9	0	0	C: 'tool,' 'thing,' 'rifle/gun,' 'war,' 'fight'
97. 'window'	**44.6**	**36.9**	7.7	18	3.5	7.1	0	7.6	0.3	1	0.8	16.0	0	0	C: 'door,' 'hole/opening,' 'see,' 'light/bright,' 'house' A: 'mirror' (North America)
BODY PARTS AND BODY FLUIDS															
98. 'Adam's Apple'	**53.4**	**37.0**	20.0	NA	18.8	0.7	0	3.3	0	0	1.2	28.5	0	0	C: 'throat/neck,' 'larynx'
99. 'ankle'	**46.1**	**41.7**	6.4	50	11.8	0	0	1.8	0	0	2.2	28.4	0	0	C: 'foot/leg,' 'joint/wrist,' 'eye,' 'neck/throat'

App. D (cont'd)

100. 'beard'	**51.8**	**33.5**	19.9	14	23.4	0	0	1.1	0	3	0.6	15.9	0	0	C: 'hair,' 'chin,' 'mouth,' 'whiskers,' 'fur' A: 'feather' (Mainland Southeast Asia and Oceania)
101. 'bladder'	**67.2**	**47.3**	19.9	NA	0	31.6	0	0	0	0	0.7	12.3	10.2	0.8	C: 'urine/urinate,' 'container'
102. 'blood'	**15.4**	**2.4**	13.0	0	0	0.7	1.4	5.3	0	0	0.2	1.5	0	0	
103. 'bone'	**21.7**	**0.7**	20.3	6	0.7	0	0	0	0	5.4	1.8	10.3	0.7	0.1	A: 'strong/strength' (New Guinea)
104. 'brain'	**33.8**	**20.2**	13.7	16	4.9	3.4	0	0.8	0	0	2.5	22.2	0	0	C: 'head,' 'marrow'
105. 'breast'	**57.8**	**5.2**	53.0	6	10.5	4.8	0	0.7	16.6	0	0.2	0.6	5	8.7	C: 'milk,' 'udder/teat,' 'nipple/teat,' 'suck,' 'chest' A: 'bay' (Europe), 'Burton's legless lizard' (Australia)
106. 'buttocks'	**55.4**	**17.5**	37.4	24	22.0	0.4	0	0	1	0	0.8	17.9	0	0	C: 'bottom/base,' 'back/behind,' 'hip/haunch,' 'anus,' 'rump'
107. 'calf'	**43.4**	**34.0**	10.5	40	11.7	0	0	0.9	0	0.4	2.2	28.5	0	0	C: 'leg/foot,' 'belly/stomach'
108. 'cheek'	**26.3**	**10.2**	16.5	16	17.1	0	0	2.4	0.7	0	0.2	4.2	0	0	C: 'face,' 'jaw'
109. 'chin'	**44.2**	**12.8**	33.0	24	40.5	0	0	0	0	0	0.1	3.8	0	0	C: 'jaw,' 'beard'
110. 'eyeball'	**86.9**	**81.0**	7.5	NA	9.4	0	0	8.1	0	0	3.7	65.4	0	0	C: 'eye,' 'seed,' 'egg,' 'child,' 'pupil,' 'fruit,' 'grain' A: 'round object' (Eastern North America)
111. 'eyebrow'	**44.6**	**35.3**	9.3	36	15.7	0	0	0	0	0	0.6	7.2	2.6	0.4	C: 'eye,' 'hair,' 'eyelash,' 'fur,' 'feather'
112. 'eyelash'	**61.4**	**53.6**	8.0	60	28.2	0	0	0	0	0.8	0.3	8	0.8	0.1	C: 'eye,' 'hair,' 'eyebrow,' 'fur,' 'feather,' 'eyelid'
113. 'eyelid'	**73.7**	**68.9**	5.3	80	28.2	1.2	0	0	0	0	0.7	9.4	11.4	1.2	C: 'eye,' 'skin,' 'peel/rind/shell,' 'bark,' 'leather/hide,' 'eyelash,' 'cover/lid'
114. 'finger'	**59.6**	**22.1**	37.5	36	14.5	1.8	0	4.2	0	0	1.7	33.9	0.8	0	C: 'hand,' 'toe,' 'arm,' 'claw/paw/forefoot'

App. D (cont'd)

115. 'fingernail'	**69.2**	**12.8**	56.1	26	1.3	0	0	0	0	0	35.7	15.1	31.3	2.1	C: 'claw/talon,' 'toenail,' 'hoof,' 'finger,' 'hand'
116. 'guts'	**47.7**	**13.0**	34.9	28	3.9	0	0	0	0	0	1.5	5.9	0	0	C: 'innards/bowels,' 'belly/stomach,' 'faeces' A: 'end' (Eurasia)
117. 'heart'	**48.0**	**11.5**	37.3	8	2.7	1.7	0	0.3	0	0	4.8	22.3	0.3	0	C: 'feel/think,' 'center/middle,' 'soul/spirit'
118. 'jaw'	**57.6**	**20.4**	37.3	26	49.6	0	0	0	0.4	0	0.1	4.1	0	0	C: 'chin,' 'cheek'
119. 'kidney'	**24.6**	**17.1**	7.5	18	2.3	0	0	0	0	0	5.9	13.5	0	0	
120. 'lip'	**47.4**	**23.5**	24.6	28	14.6	0.7	0	0.7	0.5	0	0.6	7.1	1.8	0.3	C: 'mouth,' 'edge,' 'skin'
121. 'liver'	**13.4**	**1.5**	11.9	6	0	0	0	0	0	0	NA	4.7	0	0	
122. 'lungs'	**25.4**	**16.5**	8.9	18	1.4	1.4	0	4.9	0	0	0.7	5.8	0	0	C: 'liver'
123. 'milk'	**54.7**	**23.9**	30.8	22	0	3.7	0	0	45.2	0	0	2.4	0	0	C: 'breast,' 'water/liquid/juice,' 'udder/teat,' 'nipple/teat' A: 'Burton's legless lizard' (Australia)
124. 'mouth'	**48.8**	**5.9**	42.9	4	7.6	6.7	0	0	0.9	0	0.4	1.4	5.2	3.8	C: 'opening,' 'word/language/speech,' 'beak,' 'lip,' 'edge/tip,' 'door/entrance'
125. 'phlegm'	**51.3**	**12.2**	39.1	NA	0.7	0.3	0	0	12.2	0	2.3	30.0	0	0	C: 'saliva/spittle,' 'cold/flu,' 'snot,' 'cough'
126. 'navel'	**28.1**	**8.3**	20.5	4	16.4	0.7	0	0	0	0	0.3	5.5	0	0	C: 'umbilical cord'
127. 'neck'	**39.9**	**2.9**	36.9	8	27	0	0	0	3.3	0	0.2	4.7	0	0	C: 'throat' A: 'nape' (South America)
128. 'nipple'	**74.2**	**58.2**	17.2	58	7.2	2.4	0	14.2	2.1	0	1.0	26.4	0.4	0	C: 'breast,' 'point/tip/end,' 'milk,' 'udder,' 'eye,' 'head'
129. 'nostrils'	**78.7**	**76.8**	2.9	88	53.7	1.9	0	0.5	0	1	0.2	8.9	1	0.1	C: 'nose,' 'hole/opening'
130. 'pupil'	**75.4**	**65.7**	9.6	NA	5.4	0	0	22.1	0	0	1.5	40.1	0	0	C: 'eye,' 'black,' 'child/son/daughter,' 'eyeball,' 'seed/grain,' 'iris,' 'small'

App. D (cont'd)

131. 'pus'	**23.4**	**5.6**	17.8	8	0.9	0	0	0.9	5.2	0	1.4	9.5	0	0	
132. 'rib'	**33.6**	**23.1**	12.0	40	27.4	0	0	0	0	0	0.2	4.5	0	0	C: 'side/flank,' 'bone,' 'chest/thorax/ribcage'
133. 'saliva/spittle	**34.3**	**18.0**	16.4	0	8.9	2.5	0	0	0	0	1.4	16.2	0	0	C: 'water/liquid/juice,' 'phlegm/sputum,' 'mouth' A: 'foam' (Mainland Southeast Asia and Oceania)
134. 'scar'	**33.2**	**12.3**	21.7	30	0	0	0	0	12.8	1.9	0.7	9.6	0	0	C: 'wound/sore,' 'mark'
135. 'skin'	**73.4**	**3.1**	70.0	16	2.3	0	0	0	17.7	0	1.3	0	26.0	NA	C: 'leather/hide,' 'bark,' 'rind/peel,' 'shell,' 'surface/cover,' 'husk/chaff'
136. 'snot'	**37.4**	**15.8**	21.6	18	1.4	0	0	0.9	12.8	0	1.0	14.3	0	0	C: 'nose,' 'cold/flu,' 'phlegm/mucus'
137. 'semen'	**46.7**	**26.8**	21.5	NA	0	0	0	2.3	7.6	0	1.9	17.7	0.8	0	C: 'water/juice/liquid,' 'seed,' 'penis,' 'pus,' 'child'
138. 'stomach/belly'	**32.5**	**9.2**	23.6	6	4.5	1.3	0	0	0	0	0.8	2.2	2.5	1.2	C: 'guts/innards,' 'womb,' 'heart'
139. 'sweat'	**17.7**	**7.7**	10.1	NA	0	0	0	0.9	10.5	0	0.3	3.2	0	0	C: 'heat/warmth'
140. 'tear'	**41.6**	**37.8**	3.8	36	0	0	0	0.2	33.5	0.4	0	0.4	0.8	2.0	C: 'water/liquid/juice,' 'eye,' 'cry'
141. 'tendon'	**70.2**	**5.7**	64.5	8	4.1	0	0	0	0.5	0	0	0	0	NA	C: 'vein/artery,' 'nerve,' 'muscle,' 'thread/twine/fibre,' 'root'
142. 'testicle'	**47.4**	**21.4**	26.0	26	8.6	0	0	1.8	0	0	1.1	11.5	0.4	0	C: 'egg,' 'scrotum,' 'seed'
143. 'tongue'	**20.6**	**3.1**	17.5	10	0.3	10.9	0	0	0	0	0.6	6.3	0	0	C: 'language/dialect/speech/word'
144. 'tooth'	**22.5**	**3.1**	19.4	6	2.1	1.7	0	0.3	0	0	3.4	13.9	0	0	
145. 'urine'	**12.0**	**5.7**	7.7	NA	0	0	0	0	0	0	NA	2.7	0	0	

App. D (cont'd)

146. 'uvula'	**60.8**	**45.8**	15.0	NA	18.8	2.5	0	0	0	0	1.8	38.8	0	0	C: 'tongue,' 'throat,' 'child/son,' 'vagina/clitoris,' 'tonsil'
147. 'vein'	**65.0**	**20.0**	46.8	41	4.6	0	0	0	0	0.8	3.3	8.7	9.2	1.1	C: 'tendon,' 'nerve,' 'blood,' 'way/street,' 'muscle'
148. 'womb'	**67.6**	**48.6**	23.3	42	6	15.5	0	1.3	0	0	1.3	1.1	28.8	27.0	C: 'child/son/daughter,' 'stomach,' 'house'
149. 'wrinkle'	**22.9**	**8.1**	14.8	NA	2.8	0	0	1.8	0	0	3.4	15.5	0	0	C: 'crease/fold/pleat,' 'ripple'
PHASES OF THE DAY AND MISCELLANEA															
150. 'dawn'	**51.0**	**33.8**	17.1	48	0	0.2	15.9	14.5	0	0	0.3	8.6	0	0	C: 'light/light up,' 'day,' 'morning,' 'sunrise'
151. 'day'	**53.9**	**4.5**	49.4	18	0	0	22.6	3.7	0	3.1	0	0	0	NA	C: 'sun,' 'daylight/light,' 'time'
152. 'dusk'	**55.6**	**31.6**	24.8	NA	0	1.9	18.2	15.4	0	0	0.2	8.1	0	0	C: 'dark/darkness,' 'twilight,' 'night,' 'afternoon,' 'evening,' 'sunset,' 'sun'
153. 'night'	**26.7**	**7.0**	20.6	14	0	0	7.5	17.1	0	0	0	0	0	NA	C: 'dark/get dark,' 'evening' A: 'sleep' (Australia)
154. 'noon'	**65.8**	**58.3**	9.2	72	0.4	0	12.3	1.9	0	37	0	1.7	0	0	C: 'day,' 'middle,' 'sun'
155. 'sunrise'	**65.6**	**52.3**	13.3	NA	0.6	0	19.1	3	0	57.7	2.5	2.6	0	0	C: 'sun,' 'come out/go out/emerge,' 'dawn,' 'appear,' 'east,' 'rise,' 'light/daylight' A: 'sprout' (Africa)
156. 'sunset'	**67.2**	**50.7**	17.7	NA	0	0	5	10.5	0	0	3.2	38.3	0	0	C: 'sun,' 'time,' 'sink,' 'enter,' 'fall/drop,' 'dusk,' descend/go down/lower,' 'day'
157. 'man (human being)'	**35.1**	**8.9**	26.2	14	0.4	0	4.5	4.6	0	0	0	30.1	0	0	C: 'member of ethnic group,' 'man'
158. 'Saturday'	**33.9**	**33.8**	0.2	32	2.7	0	0	3.8	0	3.8	0.2	0	0	NA	C: 'day,' 'six' A: 'younger sibling' (Eastern North America), 'Saturn,' 'unique' (Europe and Eurasia)

App. D (cont'd)

159. 'virgin'	**53.3**	**53.0**	2.7	NA	0	3.2	8	1.8	0	22.9	0	7.1	0	0	C: 'girl/daughter,' 'unmarried girl,' 'man,' 'NEG,' 'closed/shut'
160. 'widow'	**32.1**	**26.9**	5.2	52	0	0	0	0	0	20.5	0.1	0.4	0	0	C: 'woman,' 'husband,' 'die/dead'

Legend

MSC = Modified Simplicity Score from Haspelmath and Tadmor (2009c), SC = Spatial Contiguity, FC = Functional Contiguity, TC = Temporal Contiguity, PC = Perceptual Contiguity, PRC = Provenience Contiguity, CC = Configurational Contiguity, CS-Ratio = Contiguity/Similarity-Ratio, PS = Perceptual Similarity, FS = Functional Similarity, PF-Ratio = Perceptual Similarity/Functional Similarity-Ratio

* Number of languages for which data is available ≤ 2.

† Averaged from values for 'lightning' and 'bolt of lightning' (cf. § 5.2.1.).

‡ 'shore'

Appendix E

Lexico-Semantic Associations

Introduction and notes on data presentation

If motivated terms are coined or semantic extensions are institutionalized for whatever reasons, they necessarily bear some lexico-semantic associations by the very definition of the concept of lexical motivation, and there is little reason to believe that there is a correlation between the number of terms and their semantic structure.

Thus, however many motivated terms are found in an individual language, speakers in each case, for each meaning to be expressed, have to make a selection out of possible semantic associations. This section surveys the linguistic treatment of each meaning on the meaning list of the present study, and provides short accounts of the cross-linguistic variation found. In this sense, the model of this section is the still unrivalled work of Buck (1949) on Indo-European, and reference will be made throughout to it to compare the cross-linguistic findings with those found in this particular language family. However, there are also important differences from Buck's pioneering work: first, of course, the evidence presented here is based on a world-wide typological sample from languages of diverse genealogical affiliation, and is not restricted to one language family. Second, in contrast to Buck, the present account is synchronically oriented, as opposed to Buck's interweaving synchronic observations on lexico-semantic associations with diachronic developments. Third, the style of presentation is different, and the data provided here are more detailed than the discussion of the meanings in Buck: there will be ample data from the sampled languages to illustrate the observed patterns, and precise numbers as to their strength will be provided. In this sense, the chapter is a first step to answer the voices calling for a systematic cross-linguistic investigation of lexico-semantic associations discussed in chapter 2.

However, apart from a mere account of the data, there is more merit to such a discussion: Blank (2003) argues that Tagliavini's (1949) pioneering study of associations in terms for the 'pupil of the eye' allows to predict that it is highly likely that, if a complex term for this meaning should be coined anywhere, the associations will be drawn from the list compiled by Tagliavini. This, while at the same time extending such comparative overviews beyond a single concept, is another value of these data.

In order to make the presentation of the data maximally useful for further research and to make it readable independently from the framework developed in the present study, terminological peculiarities are kept to a minimum. However, sometimes it is useful to be able to have recourse to a number of terminological conventions to talk succinctly about the phenomena encountered. These require little to no additional theoretical assumptions, and thus do not distort the data in terms of a particular theoretical framework. Similarity and contiguity are, purely descriptively, taken to be the fundamen-

tal semantic relations underlying METAPHOR and METONYMY. respectively. For instance, Cahuilla uses the same term for 'finger' and 'hand' (in more technical parlance that will be employed in the following, Cahuilla COLEXIFIES these meanings), while Aguaruna uses the same term for (colexifies) 'finger' and 'branch.' Now, the meanings 'finger' and 'hand' stand in a relationship of contiguity to each other (they are spatially adjacent to each other, or, since indeed the fingers are part of the hand, 'finger' is a meronym of 'hand'), while this cannot be said of 'finger' and 'branch': rather, they are similar to each other in their longish shape and the fact that they protrude from a larger entity, the (rest of the) hand and the trunk of a tree respectively. In chapter 3, the semantic relationships of contiguity and similarity are defined in terms of test frames, and the categorization in the following relies on these as the criterion as to which fundamental semantic relation should be posited; however, since these test frames are explicitly designed to capture intuitions as to which relation is present it is not strictly speaking necessary for the reader to bear in mind in detail how these tests work in order to get an overview of the linguistic treatment of the concepts of interest to him/her. In analyzable terms which contain more than one lexical element, CONTIGUITY ANCHORING describes the semantics of one of the two constituents which ties back a metaphorical transfer accomplished by the other element to the semantic domain of the target concept. For instance, in Katcha *ɔe mɔ mbɔrɔ* 'nostril,' literally 'eye of nose,' there is a metaphorical transfer from *ɔe* 'eye,' the head of the complex term, to 'nostril,' while *mbɔrɔ* 'nose' is in a relation of contiguity to the target concept 'nostril' and anchors the metaphorical term in the domain of the target concept.

Some notes on presentation style are in order: each section begins with an overview of the most common lexico-semantic associations related to the meaning under discussion, in descending order of their occurence in the languages of the sample. Further, there are percent values giving an idea of whether for this meaning, motivated terms are dominantly contiguity- or similarity-based (here, reported values often do not add up to the total number of motivated terms, which is either due to several possible analyses of the semantic association(s) or it being unclear), and whether they are typically by morphological analyzability or more often by colexification. Also, for each of the meanings, the number of languages for which an equivalent could be retrieved from the consulted sources is stated. In the body of the text, languages betraying a given recurrent lexico-semantic association are named first according to the macro area from Dryer (2005) they belong to, and are, within areas, listed alphabetically. Associations only found in one particular language of the sample are typically listed in a separate section that follows the one discussing recurring associations, except if they relate in some way to one of the associations discussed earlier in the text. If additional meanings mentioned in the consulted sources (other than ones that are so close to the item on the meaning list or one of the recurrent associations which is already discussed that mentioning them specifically seems redundant) are omitted from the discussion, this is indicated by phrases such as "inter alia." This is mostly done for languages in which the relevant terms have a wide range of meanings none of which seem to be standing in any obvious semantic relation to one another. For this section in particular, it must be emphasized that formulations such as "the term x in language y also means z" or "colexifies z" does not necessarily entail the claim

that the meanings are related semantically. This is true generally, but should be borne in mind particularly when it comes to statements pertaining to Mandarin Chinese, where phonological changes discussed in § 5.4.2.3.2. lead to the collapse of a large number of erstwhile distinct lexical items.

When longer passages or longer glosses anywhere in the following are quoted directly from a consulted source, they are given in double quotation marks (this does not exclude the possibility that shorter glosses given in single quotes coincide with the gloss in the source). In the case that phonological processes alter the shape of the constituent morphemes of an analyzable term on the surface, the (inferred) morphological analysis is given in square brackets after the relevant object language term, as in Ket *destul* /dēs-d-ūl/, literally 'eye-POSS-water' and meaning 'tear.' Otherwise, the surface form in italics is segmented directly in order to save space. Grammatical material in glosses is, as in this example, printed in small caps and usually abbreviated; a list of these abbreviations can be found in the front matter. Possibly existing analyzable terms of the redundant type (as when the simplex *ange* in Kosarek Yale denotes both 'umbilical cord' and 'navel,' but there also is the complex term *ange lom*, with *lom* meaning 'hole, valley,' to single out the meaning 'navel') are usually not mentioned.

Discussions of individual meanings are referred to as "sections" along with their respective numbers in cross-references within this appendix, while cross-references to chapters in the main text are indicated by a paragraph sign (§).

The proxies, if any, that were accepted for a given meaning on the list are identical generally to those in Buck (1949), unless otherwise noted. For instance, as in Buck (1949), no attempt to distinguish between (bigger) 'lake' and (smaller) 'pond' is made. As noted in chapter three, a given term may exhibit more than one lexico-semantic association, either by virtue of being of the lexical type, or by being analyzable, but at the same time also colexifying more than one meaning. The latter fact in particular is a problem for a smooth and readable discussion of the associations found in terms for a given meaning, since it would require discussion constantly to jump back and forth between different associations. For instance, the Fijian word for 'mirror,' *i iloilo*, contains the instrument nominalizer *i* and a reduplication of the verb *ilo* 'to look at.' However, at the same time, it also colexifies 'glass' in general, a pattern that other languages exhibit as well, and thus Fijian has to be mentioned twice in the discussion, first in discussing terms derived from verbs meaning 'to look,' and second in terms colexifying 'mirror' with 'glass.' Since it is impossible to always do so without proliferating cross-references within each individual section ad absurdum, it is not always indicated which term bears which multiple semantic associations. This can be inferred simply from languages being mentioned twice, except if a language has several terms for the meaning and they exhibit different lexico-semantic associations. Conversely, however, it may be that languages have synonyms for the meanings in question and different ones bear different associations. That is, if a language is mentioned several times in each section, this does not necessarily always entail that it is always the same term bearing all associations.

Moreover, not all patterns of colexification are mentioned for constituents in complex terms. For instance, terms for 'tear' often consist of elements meaning 'eye' and

'water,' with 'eye' having secondary readings (see section 140 and § 6.2.3.1.) which are not mentioned since clearly it is the core meaning 'eye' that is relevant in the conceptualization of 'tear.' Any other way of proceeding would yield a resulting discussion that is highly cumbersome to read. This policy is departed from in cases where it seems unclear which sense of the constituent is relevant for the conceptualization of the complex term, in which case different meanings of constituents are separated by slashes in the gloss.

A further issue is ambiguitites arising due to strictly speaking conflicting information in two-way dictionaries. For instance, Nez Perce *hím̓* is stated in the Nez Perce-English section of the consulted source to mean 'mouth; mouth of river, cave,' but when one seeks to identify the Nez Perce term for 'lip' from the English-Nez Perce section, one also encounters *hím̓* as one of the equivalents. In this and other such cases, all information was taken into account, that is, Nez Perce is mentioned as a language colexifying 'lip' and 'mouth' and is mentioned as such in both relevant sections, in spite of being not explicitly glossed as 'lip' in the dictionary.

1. The Animal

Representation: 83%
Motivated: 43.1%
Thereof Analyzable: 26.8% Thereof Colexifying: 18.3%
Thereof by Contiguity: 32.8% Thereof by Similarity: 1.6%
Recurrent associated meanings: thing, meat, life/be alive, bird, livestock, cattle, move, pig, dog, insect, forest/wood, kill, move on ground, brute, person, land

A common association, by contiguity, is that with 'meat' or 'flesh,' since obviously animals are the source of meat to be consumed by humans. The association is mostly realized by colexification and occurs in Efik, Hausa, Ngambay, Noni, Yoruba, Gurindji, Yir Yoront, Abzakh Adyghe, Sora, and Yanomámi, and by noun class alternation in Swahili.

Another common pattern is to have terms for 'animal' which also at the same time mean 'thing' very generally. This is found by colexification in Buli, Kwoma, Arabela, Chayahuita, Lengua, Wayampi (where the relevant term can also mean 'kind of' and 'luggage'), and Samoan. In Lengua, as well as in Khalkha, the term may also refer to a 'person' and in Buli also to a 'figure, unrecognizable person;' compare also Copainalá Zoque *copʌn* 'animal' and *pʌn* 'person.' Furthermore, in Katcha, there are the complex terms *nimo mo tile* 'thing of forest' and *nimo mo di* 'thing of house' to refer to wild as opposed to domestic animals respectively (for the former term, compare also Bora *bájú-e-jpi* 'mountain/forest-belong.to-CL.M.S,' and Gravelle's 2004: 375 statement that a similar term is found in Meyah; the distinction between domestic and wild animals is also made in Rendille, Oneida, and Aymara; for Cushitic specifically see also discussion in Sasse 2002). Similarly, Itzaj has *b'a'al-che'* 'thing-wood/tree' for 'animal, fauna.' Still further, Bwe Karen has the somewhat unclear term *dɛ-pho=dɛ-wɛ* 'thing-child/little.one=thing-bug' glossed as 'dumb creatures, animals' (compare colexification of 'animal' and 'bug' in Highland Chontal and of 'animal' and 'insect' in Nivkh, Kildin Saami, and Hawaiian). There is a semianalyzable term with the identifiable constituent meaning 'thing' in Sko.

Mandarin has *dong4-wu4* 'move-object/being,' and Japanese and Vietnamese, due to Chinese influence, *dō-butsu* 'move-thing' and *động mật* 'move object' respectively. Moreover, Hani, a language also spoken in Southeast Asia, has *nivzeig ~ nilzeig*, maybe related to *niv* 'moving; action' (though note also *nil*, meaning 'the outside, wild' inter alia), Nuuchahnulth has *saštuup ~ saxtuup* /sa-tuˑp/ 'crawl.on.all.fours-creature,' (with *tuˑp* also glossed as 'thing' generally as well as 'kind, sort;' compare Chukchi *ɣennik* which is perhaps related to *-ŋərtə-* 'four'). Similarly, Blackfoot has *iksowá'pomaahkaa* /iksow-á'p-omááhkaa/ 'at.ground.level-about/around-move.along.on.foot,' and Lakhota *wamákhašką* is literally translatable as "those moving about on earth."

Returning to associations with 'thing,' Khoekhoe has *ûitsama tsaba xūn*, with *ûitsama* meaning 'living' (derived from *ûi* 'to escape, to escape death, be alive'), and *xūn* referring to 'possessions' and 'livestock.' Both associations recur: 'livestock' is also colexifed in Rendille, Basque, and Yay, and relevant terms in Gurindji, Bezhta, and Samoan also denote 'cattle' specifically. As for the association with 'life' or 'be alive,' there are derived terms in Muna and perhaps in Greek, Kiliwa has *t-kw+ipaa-y* 'OBJ-WH+be.alive-ATT,' Nez Perce *waqíˑswitin* /waqíˑswit-iˑn/ 'life-with,' and Carrier *rhenna* 'large animal' is a verbal noun derived from *rhesna* 'to be alive.' Lenakel combines the association with the already familiar pattern relating 'animal' and 'thing:' *nar amíuh* is analyzable as 'thing alive,' and Piro *giwekachri yotaljetachri* relates again to the association with 'movement': it contains elements meaning 'to be alive' and 'to move.' Yoruba *ẹlẹ́mìí* is analyzable as /ẹlẹ-ẹ̀mí/ 'owner-breath/life' (there is also the term *ẹ̀-dá* 'NMLZ-to.create' for 'creature, creation'). Further, Central Yup'ik *ungungssiq* contains the base *unguva* 'life,' in Samoan, there is the redundant analyzable term *meaola*, with *mea* meaning 'thing, animal' (and also 'genitals' and 'place' generally inter alia), and *ola* 'living,' and there are (probably) further terms that are diachronically related to verbs meaning 'to live' in Khalkha (here, the relevant term seems to have further connections to terms related to 'breathing'), and Kolyma Yukaghir.

Upper Chehalis *χəs=áy'=tm'š* is (semi)analyzable as 'bad=??=land,' and Central Yup'ik has a term for land animals, *nuna-miutaq*, which is analyzable as 'land-one.whose.proper.place.is.' A pattern peculiar to New Guinea is to have dvandva compounds expressing the concept 'animal,' with the constituents denoting particular animal species: thus, Kyaka has *suwua-pe saa-pe mena-pe pyasingi* 'dog-ASSOC furred.animals-ASSOC pig-ASSOC mixed.assorted.group,' Sentani *obo-joku* 'pig-dog,' and Takia *bor-goun* likewise 'pig-dog,' but here the compound has reference to 'domestic animals' specifically. Hawaiian has *holoholo-na* 'walk-NMLZ,' and there is a semianalyzable term featuring a constituent meaning 'walk' in Great Andamanese. Haida has *gina ti7araa* 'creature/thing be.killable' for 'land mammals' specifically, and the Oneida term for 'wild animals,' *kutilyoʔshúha*, is likewise derived from a verb meaning 'to beat, kill.' Lesser Antillean Creole French and Basque colexify 'animal' with 'brute.' Kwoma *boboy* also means 'plant,' and the association might also be present in Abipón, but this is considered unsure in the consulted source.

Berik and Biloxi make a distinction between 'female animal' and 'male animal' (the Berik term for 'wild animal,' *giri*, also means 'deep river'). Upper Chehalis, Highland Chontal, Huambisa, Jarawara, Fijian, Kapingamarangi, Rotuman, and Samoan colexify

'animal' with 'bird,' that is, a lower rank in ethnobiological taxonomies, and 'animal' is colexified with specific animals on the generic level in Dadibi ('cuscus'), Sora ('pig'), Kolyma Yukaghir ('elk'), Pawnee ('deer'), and Central Yup'ik ('bear').

Other associations include: Buli *dung* also means 'to press down, press out' inter alia, and Efik *u'nam* might be related to *nam* with the basic meaning 'to do, make' (considered unsure in the consulted source). Ngambay *da* also means "to assemble something" and also denotes a "kind of tree, bark strings to attach the roof of a house with." Noni *nyam* might consist of the verb *yam* 'to suck' and a prefixal noun class marker. The Anggor term *nine-hondɨ* is analyzable as 'fur/feather-mother.' Khalkha *aduɣusu(n)* is derived from *aduɣu(n)* 'herd of horses, horse' by means of the suffix *-sun* the function of which is, according to Poppe (1954: 44), "to form nouns of which the meaning is usually the same as that of the primary word." Ket has *assel* /ēs-sèl/ 'wild reindeer,' which is likely an instance of a so-called markedness reversal (Berlin 1972, Witkowski and Brown 1983). The Cahuilla term *ʔíʔihiŋaviš* is likely to contain *-ʔi* 'leg, foot, footstep, track.' Kashaya *šiʔbaši* contains *šiʔba* 'body,' and Kiowa has a term for domestic animal, *yʜtbʜ-dou'*, which is analyzable as 'go.live.with-have.' Quileute *ʔixʷáɬso* appears to be related lexically to *ʔixʷáɬsil* 'to hunt,' and Tuscarora *yuʔtikę́hraʼt* contains the roots *-(ę)ʔtikęh(r)-* 'mind' and *-aT-* 'stand.' The Yana term *mooyau(na)* consists of *ma-* 'to eat' and the nominalizer *-yau(na)*, and Yuki *he´lik´ke* appears to contain *lik* 'to swallow.' Embera *ãrĩmárã* also means 'inhuman' and 'cruel,' Sáliba *omaĩdi* also 'heart,' and Yanomámi *yaro* can in some contexts also refer to an 'enemy warrior.' Finally, Kapingamarangi *manu* is also the name of a constellation involving the star Sirius.

2. The Ashes

Representation: 96%
Motivated: 38.5%
Thereof Analyzable: 17.4% Thereof Colexifying: 21.0%
Thereof by Contiguity: 5.3% Thereof by Similarity: 18.4%
Recurrent associated meanings: dust, fire, powder, embers, coal, faeces, dirt, gray, kitchen/fireplace, wood, lime, sand, black, soot, blue, soil, flour, burn, feather

Intra-domain associations with other fire-related terms, namely 'coal' and 'embers,' are frequent for 'ashes' (or 'cinders'), see also sections 13 and 19. Five sampled languages, Burarra, Kwoma, Yir Yoront, Khalkha, and Sedang colexify 'ashes' and 'coal,' and there are derived terms in Wintu (colexifying 'soot, soot carried by the wind;' for the association with 'soot,' compare also Hawaiian *paʻu ahi* 'soot fire,' denoting 'soot' next to "black cindery sand or ash") and Great Andamanese. In a parallel fashion, Kosarek Yale, Khalkha, Carrier, Kiliwa (by the complex term *kw+pal* 'PERF+hot'), Cubeo, and Ancash Quechua colexify 'embers,' while Aymara has *nina sank'a* and *nina japu* 'fire/embers embers.' More generally, mirroring an association also diachronically attested in Indo-European (Buck 1949: 73-74), many sampled languages employ the same term for 'ashes' and 'dust,' namely Efik (where the relevant terms also can refer to a "dimness of vision as if a mist were before the eyes" and a kind of spearmint), Berik, Burarra (where the relevant term more

precisely means "little particles, as ashes and sand mixed where a fire has been burning" as well as 'dust'), Lavukaleve, Mali, Nunggubuyu, Rotokas (colexifying also 'wood chips'), Basque, Upper Chehalis ('cold ash' specifically), Chukchi, Biloxi, Carrier, Highland Chontal, Lesser Antillean Creole French, Lake Miwok, Oneida, Tuscarora, Wintu (where *bukul* contains *buk* 'dark'), Yuki (where *poʔoṭel* contains *poʔ-* 'burn' and *poṭ-* 'gray'), Cavineña (where the relevant term means 'dust cloud' more specifically), Hani, and Bwe Karen. Alternatively, some languages have complex terms for 'ashes' based on 'dust,' mostly with 'fire' acting as a contiguity anchor, as in Northern Yana *ɬabʔlaawi-ʔau(na)* 'dust-fire.' Such terms are also found in Carrier, Nez Perce, and Wichí; alternatively, in Pawnee, 'ashes' is *karaak-tuuhcuˀ*, analyzable as /itkaar-haak-tuuhc-uˀ/ 'dust-wood-??-NOM' and in Hawaiian *lepo uli*, analyzable as 'dust/dirt/soil/rubbish/excrement dark.color' (a semianalyzable term with the constituent 'dust' is found in Kashaya and San Lucas Quiaviní Zapotec).

For the Yuki association with 'gray,' note also that Muna *ghabu* may also refer to the "grey colour of certain fruits (indicating old age)" (cf. also *ghabu ~ ghadu* 'enormous, huge'), as well as that Tuscarora *uʔkę́hreh* colexifies 'ashes,' 'dust,' and 'gray,' Huambisa 'ashes,' 'gray,' and 'lead,' and Hawaiian *lehu* means 'gray,' but is used more specifically to describe the color of chickens inter alia. On a related note, Yanomámi *ushi pë* appears to consist of *ushi* 'dark blue, dark violet, bruise, ripe' and the quantal classifier *pë* (see § 4.4.1.1.), while Yay *taw*^{5} colexifies 'ashes' with "blue, ash-color." Colexification with 'dirt' is attested for Oneida and Wintu, here by the same semianalyzable term mentioned above (in both languages, 'soil' is also colexified, and Yanomámi colexifies a specific type of dark soil with 'ashes'), and complex terms betraying this semantic association are found in Yir Yoront (*thum-nhur* 'fire-dirt'), Wappo (*hél-pi-poo̲l* 'fire-from-dirt') and in Hawaiian, as seen above. Still more generally, colexification with 'powder' occurs in Ngambay, Kyaka ('white powder' specifically), Basque, Chukchi, Cheyenne, Highland Chontal, and Bora, and a realization of the association by morphological complexity is found in Piro (*tšit̲šu̲-pahi* 'hearth-fire/firewood-powder') and Manange (*ɜmje-pʰra* 'fire-powder'), while there is a semianalyzable term in San Lucas Quiaviní Zapotec. More specifically, Hausa colexifies 'ashes' with 'pounded flour' in particular (compare Bororo *joru-gudu ~ ru-gudu* 'fire-flour'). These semantic associations mirror closely those reported for 'dust' in section 16 (see also Buck 1949: 74 for Indo-European evidence).

Metaphor-driven associations linked with 'ashes' specifically, but usually not with 'dirt' (with the exception of the evidence from Hawaiian already discussed above) are that with 'faeces' or 'excrement,' occurring by colexification in Kwoma (the term also colexifies 'flower') and by analyzable terms of the structure 'fire-faeces' in Cayapa, Bislama, and Takia, while Itzaj has *taˈan* /taˈ-Vn/ 'excrement-DERIV' (note also the somewhat unclear case of Kiowa *sɑ̄ʾ-pʿn̥*, perhaps 'excrement-cloud/sky'). The Itzaj term betraying this connection in addition colexifies also 'lime,' while in Copainalá Zoque *cuy-jam* is analyzable as 'tree/wood-lime.' Outside Mesoamerica, Tetun has *ahu-metan* 'lime-black' (for which in turn compare Santiago Mexquititlan Otomí *ˈbo-spi* 'black-fire' and Kyaka *keyihapa now*, where *keyihapa* is 'black' and *now* is a term for color and coloring earth pigments), and Fijian colexifies 'ashes' with 'slacked lime' specifically (the relevant term *dravu* means 'to rub the head with ashes' as a verb). There is an association with 'feather' found in Tsafiki

(*nin fu* 'fire feather/body.hair') and Great Andamanese (*châpa-l'ig-pîd* 'firewood-??-feather').

Further complex terms in which 'fire' acts as a contiguity anchor include Lesser Antillean Creole French *sann dife* 'sand fire' (compare colexification of 'ashes' and 'sand' in Cayapa and Hawaiian), and Sedang *pló on* 'fermenting.agent fire.' Semianalyzable terms with 'fire' are found in addition in Mbum, Toaripi, Kosarek Yale, Nez Perce, and Pipil. Sedang *pló on* 'fermenting.agent fire' also colexifies 'hearth' or 'fireplace,' and the association with 'fireplace' or 'kitchen' is mirrored by unanalyzable or semianalyzable terms in Muna, Sahu, and Toaripi (note also the apparent relationship Guaraní *tanimbu* 'ashes, remains of things that were, debris' and *tanimbupa* 'kitchen, hearth'). There are also some terms which highlight that the ashes are the remnants of a fire. These include Abzakh Adyghe *safe* /sə-efe ~ sə-afe/ 'burn-REST' (compare the association between 'ashes' and verbs meaning 'to burn' in Indo-European languages, Buck 1949: 73), Central Yup'ik *qamlleq* /qame-lleq/ 'die.down-one.that.has,' and Samoan *lefulefu*, where the reduplication base *lefu* means 'for a fire to go out' (compare again Kiliwa *kw+pal* 'PERF-hot' as well as Welsh *gweddillion* 'ashes,' which is related to *gweddill* 'remnant').

Further associations include: Hausa *ru'bushi* 'hot ashes' is also the name of a kind of sweet pastry inter alia, and *raushi* 'hot fine ash' also means, inter alia, 'softness, tenderness.' Kyaka *pee* also is a general term for a 'receptable' and specific receptables inter alia, and Chukchi *piŋ(piŋ)* also means 'falling snow.' Japanese *hai* also means 'yes' (with the term in the meaning 'ash' perhaps being borrowed from Chinese). Kwoma has *hikishebo*, colexifying 'ash' with "black paint, black pigment used to make paint" and "object burnt by fire" as well as "earth blackened by fire." Ngaanyatjarra *tjurnpa* also means 'husks,' and Rotokas *gavuta* is also glossed as 'bed of fire.' Badaga *bu:di* also means 'ashmound' as well as 'contrary,' Khalkha *coγ* also 'sparks' and, figuratively, 'glory' and 'energy' inter alia, and Welsh *ulw* also 'utterly.' Upper Chehalis *sq̓ʷə́lnš* is derived from *q̓ʷə́lí* "roast, cook, ripe, ripen' and contains the lexical affix *=nš* 'basket' (or 'ball'). Tuscarora *uhséhareh* means also 'lye,' and the first variant of Abipón *-aci* is identical formally with terms for 'tongue' and 'tear.' Cashinahua *mapu* colexifies 'ashes' with 'brain' (the tertium comparationis being the gray color of both?), 'soap,' and 'handle of a tool' inter alia. Jarawara *hasawiri/hasawiri* also means 'smoke,' and Kapingamarangi *lehu* also 'smegma.'[1] Sedang *trôi* is also the name of a 'leaf-eating ant,' and Bislama *asis faea* can also refer to 'relic, remnant' in non-standard usage.

3. The Bark

Representation: 93%
Motivated: 68%
Thereof Analyzable: 23.7% Thereof Colexifying: 44.6%
Thereof by Contiguity: 0.6% Thereof by Similarity: 67%
Recurrent associated meanings: skin, peel/rind/shell, tree/wood, hide/leather, cover, scale, husk/chaff, crust, clothing, pod, scab of wound, hard, tire, fingernail

[1] "smegna" in the source.

Clearly, the most frequent association is that with 'skin.' It occurs either by colexification or analyzable terms (with 'tree' or, less frequently, 'wood' as contiguity anchor, as in Yei *pər par* 'tree skin' or Abzakh Adyghe *pxa-ś°e* 'wood-skin,' and sometimes also, due to the colexification of 'tree' and 'wood' in many languages as reported in section 65, with both, as in Berik *ti tifin* 'tree/wood skin'). Such terms are attested in as many as 82 sampled languages. 55 languages are of the colexifying type: Bakueri, Efik, Yoruba, Anggor, Buin, Burarra (where the term also colexifies 'wrapper, outer case' and is perhaps related to a word for 'clan'), Gurindji, Kwoma, Kyaka, Lavukaleve (colexifying 'bark of coconut tree' specifically with 'skin'), Mali (colexifying also 'body'), Muna, Ngaanyatjarra, Nunggubuyu, Sahu, Sentani (where the relevant term is also the name of a palm species and a discourse particle), Sko, Southeastern and Western Tasmanian, Toaripi, Kosarek Yale, Basque, Bezhta, Biloxi, Chickasaw, Ineseño Chumash (where the relevant term means 'smooth bark' specifically), Comanche, Haida, Pipil, Xicotepec de Juárez Totonac ('skin of leg of a person' specifically), Yaqui, Copainalá Zoque, Abipón, Bororo, Carib, Cashinahua, Cayapa, Guaraní, Hupda (colexifying also 'dish, plate, food'), Jarawara, Kaingang, Maxakalí, Piro, Ancash Quechua, Tsafiki, Wayampi, Yanomámi, Bislama, Fijian, Hawaiian, Bwe Karen, Lenakel, Malagasy, White Hmong, Rotuman, and Samoan. There is a derived term in Great Andamanese, and twenty-four languages have complex terms with 'tree' and/or 'wood' acting as contiguity anchor: Mbum, Ngambay (*ngóy gìrì kake* 'skin/peanut.shell/peeling behind tree'), Baruya, Berik, Kaluli, Meyah, One, Waris (where 'skin' is colexified with 'blood'), Yei, Yir Yoront, Abzakh Adyghe, Japanese, Kolyma Yukaghir, Wappo, Yuki, Cavineña, Miskito, Sáliba, Rama, Hani, Mandarin, Takia, Tetun (where 'tree' and 'plant' are colexified), and Yay; further, Noni has a term with this structure not for 'bark' generally, but for 'skinned bark' specifically. The association is diachronically detectable also in Chukchi. Given the fact that there are 138 languages in the sample for which data are available, this means that the association occurs in almost 60 per cent of the sampled languages, and that in all regions of the world. Kiowa is a little different in that here *'æ'k'ąę* seems to be a diminutive of *k'æe* 'skin' indicated by vowel nasalization combined with an unknown second element, and Ket *iŋ-ol-t* is analyzable as 'skin.PL-covering-NMLZ.' Furthermore, Cubeo *jocʉ-curi* is analyzable as 'wood-CLASS.TEXTURED.COVER,' and in Yanomámi *hisi*, *hi* is a term for 'wood' which also acts as a classifier for plants with wooden trunk and *si* means 'cover' generally; there also is an alternative reduplicated term *sisi* for 'bark.' Note also that there are many languages with redundantly complex terms, such as Dongolese Nubian *ǧõwwɪ-ŋ-gábad* 'tree-GEN-bark.' Buli also colexifies 'bark' and "hard skin' (e.g. of a crocodile)" specifically, and there are semianalyzable terms where the identifiable constituent means 'skin' in Copainalá Zoque (where the unknown element recurs in many terms for specific trees) and Great Andamanese.

Due to colexification of 'skin' with 'leather' or 'hide' (compare section 135) a subset of these languages also betrays this association by colexification. This is the case in Kyaka, Muna, Sentani, Ket (by the analyzable term *iŋ-ol-t* 'skin.PL-covering-NMLZ'), Cahuilla, Pipil, Yaqui, Copainalá Zoque, Abipón, Bororo, Carib, Cashinahua, Guaraní, Jarawara, Piro, Ancash Quechua, Tsafiki, Yanomámi, Hawaiian, Bwe Karen, White Hmong, Rotuman, and Samoan. Complex terms with the 'tree/wood' contiguity anchor are found in Highland

Chontal, Pawnee (where the constituent denoting 'hide' also means 'peeling' generally inter alia), Wappo, Cavineña, Miskito, Sáliba, Mandarin, and Tetun. In addition, in Sedang, *kơtôu* colexifies "bark; peeling, rind, husk; egg shell; outer edge of ...," that is, a different term for 'skin' is used, and the same is true of Chayahuita.

Another quite frequent association also exhibited by the Sedang term is that with the meaning 'shell' in general and/or 'rind, peel,' as for instance of a fruit, an egg, or a mussel more specifically, occurring in fifty-nine sampled languages, by colexification in Bakueri, Buli, Hausa, Yoruba, Kyaka, Ngaanyatjarra, Mali, Muna, Abzakh Adyghe, Badaga, Greek, Khalkha (also colexifying 'bark of birch tree' specifically), Sora, Kildin Saami (here the relvant term *kērr* also means 'hard, rough,' as is the case in White Hmong; similarly, Hawaiian has a derived term, which is, however, rare), Cahuilla, Chickasaw, Ineseño Chumash, Haida, Kiliwa, Lakhota, Lesser Antillean Creole French, San Mateo del Mar Huave, Nez Perce, Pipil, Central Yup'ik, Copainalá Zoque, Aguaruna, Bororo, Cashinahua, Cayapa, Guaraní, Huambisa, Hupda, Jarawara, Kaingang, Macaguán, Maxakalí, Piro, Ancash Quechua, Tsafiki, Yanomámi, Bislama, Fijian, Hawaiian (where the term has a very broad range of meanings, including also 'scalp' and 'surface, area' inter alia), Rotuman, Sedang, Tetun, and Vietnamese, and by complex terms in Efik (*ikpök ëtü* 'skin/bark/scale/shell/husk/pod tree'), Ngambay (*ngóy gìrì kake* 'skin/peanut.shell/peeling behind tree'), San Mateo del Mar Huave (*mipang xiül* 'shell tree;' this is a different term from the colexifying one mentioned above), Itzaj (*pach che'* 'shell tree'), Yuki (*ʔol šil* 'tree skin/shell'), Copainalá Zoque (*cu'y-unaca* 'tree-skin/shell'), Arabela (*naana co* 'tree shell'), Bora (*kó-miːʔo* 'wood-SCM.hard.shell' and *úmé-he-miːʔo* 'tree-CL.tree-SCM.hard.shell), Rama (*katuuk* /kát-úk/ 'tree-skin/coat/rind'), Sáliba, where *ĩxexe* colexifies 'skin,' 'hide,' and 'shell,' while 'bark' is *nugu ĩxexe* (*nugu*, 'trees'), and Hani (*albol alhov* 'tree-skin/peel/outer.layer.of.grain'); moreover, Bezhta *yicalo* contains an element meaning 'peel.' In most of these languages, 'skin' is also colexified; exceptions are Kiliwa, where *yal* is only used for 'bark' and 'shell,' but not for 'skin,' Tetun, where *kakun* colexifies 'peel,' 'shell' and 'bark,' but not 'skin,' Pawnee, which has *raakickuusuʔ*, analyzable as /raak-(i)ckuus-uʔ/ 'tree/wood-peeling-NOM' (with *(i)ckuus-* also being capable of referring to 'hide, shell' and other similar meanings, but apparently not to 'human skin'), as well as Bezhta, where the term already mentioned is derived from an element meaning 'peel.' As for the association with 'tree, wood,' there is one language, Abipón, which colexifies 'bark' and 'tree' directly.

Furthermore, nine sampled languages colexify 'bark' with 'scale,' namely Hausa, Lavukaleve (where the relevant term also denotes the "place coconut stalk hangs from"), Bezhta, Biloxi (where the relevant term also colexifies 'horns,' 'hoofs' and 'nails on fingers and toes,' which latter association is also present in an analyzable term in Rama), Nez Perce, Central Yup'ik, Guaraní, Jarawara, and Bwe Karen, and Efik has the complex term already mentioned. In most of the above languages, the relevant terms have a rather broad semantics which, from an extensional point of view, colexify at least one other, but frequently many more, meanings figuring in this report, so that their semantics, in these cases, may also be described intensionally as 'outer surface' of anything. In fact, in Sko, Basque, Chickasaw, Comanche, Haida, Lakhota, Guaraní, Huambisa, Maxakalí, Piro, Yanomámi, Lenakel, and Bislama, the relevant terms are also explicitly glossed as meaning

'cover, covering' (in Haida also as 'surface of waves' and in Lakhota also as 'envelopes' and 'wrappings' specifically). Instructive in this context is the example given in the consulted source for Kaingang *fár: kar vỹ, fár nĩ: ka kar, nén tánh kar, ẽg mẽg kar, ẽg ke gé* "everything has *fár*: trees, plants, animals, we do, too."[2] However, Bezhta *qal* exclusively denotes 'scale' and 'bark,' while for 'skin' and other meanings alternative terms are available.

Very broad semantic range of the relevant terms is also the case for the terms in which the following associations figure, unless otherwise noted as for Bezhta above. In seven sampled languages, Yoruba, Bakueri, Anggor, Muna, Sahu, Bislama, and Sedang, 'bark' and 'husk' or 'chaff' are colexified (in Anggor also 'container'), while Efik has the complex terms already mentioned. Kyaka, Muna, Basque, Greek, Sora, Hawaiian, and Rotuman colexify 'crust' (Greek 'crust of earth' and Sora 'incrustation' specifically; the relevant Hawaiian term is that derived from 'hard' mentioned above and may also mean "cooked crisp, as pig" inter alia), and in four sampled languages, Anggor, Sko, San Lucas Quiaviní Zapotec, and Piro, 'bark' and 'cloth,' 'clothes,' or 'clothing' are colexified (in San Lucas Quiaviní Zapotec also 'rag used as diapers' is colexified, while the relevant Sko term also denotes a tree species). Abzakh Adyghe and Yoruba also colexify 'pod' ('pod' is colexified in the element meaning 'skin' and other things figuring in the complex Efik term mentioned above), while Buli, Hausa and Hawaiian in addition colexify 'bark' and 'scab of wound' (Hausa also "scurf of a scalp disease" and "bits of skin from desquamation"). Rotuman and Samoan also colexify '(rubber) tire' (Samoan in addition "kin, kinsmen" as well as "belt (for a machine)" and 'foreskin').

Other infrequent lexico-semantic patterns include: Buli *tapagi* also colexifies "board, flat piece of wood," while Khoekhoe *ǂammi* is derived from the verb *ǂam-* 'debark.' Rendille has unrelated terms for "soft inner tree bark" and "soft outer tree bark," with the former also meaning "soft membranes" and "bags of waters in pregnancy" of animals, while in Nivkh, *oɣm* also means 'root.' Burarra *mun-ngarnama* is analyzable as 'CLASS.DOMESTIC-inner.thigh,' which illustrates the widespread mapping of body-part terms onto entities of the physical world in Australia, and in Kyaka *imbu* 'hard shell, bark' may also be used to refer to "anger, annoyance" as well as a 'bowstring.' Muna *kuli* also means to "have a simple meal without side dish." Ngaanyatjarra *miri* may also refer to "people of a certain skin colour, race of people" as well as "get a shock, be shocked," *pangki* (Northern dialect) experienced semantic extension to 'orange' by synecdoche and *piilyurru* also denotes the 'carpet snake, woma python,' "the shed skin of a snake or goanna," and a "chrysalis case, shell enclosing a grub before it turns into a moth" (see Evans 1992 on such assications in languages of Australia). Sko *ró* also denotes a tree species, and Basque colexifies "exterior, outside" and 'surface' generally, as well as 'cover,' 'spread,' 'case,' 'sheet,' and 'rascal.' Chukchi colexifies 'bark' with 'edible whale skin,' and Khalkha *xoltusu(n) ~ xoltasu(n)* also denotes 'monkshood' (*aconitum variegatum*). Xicotepec de Juárez Totonac colexifies 'copal.' The Wintu term *λaplah* contains *λap* 'lying down,' *ċahay* also means 'sheath' generally and denotes the 'inner bark of maple' as well as "deerleaf, cowleaf, the fat sheath veiling of the stomach of an animal ..." specifically, and *qoˑq* also

[2] Original Portuguese translation from the source: "Tudo tem 'pele': árvores, plantas, animais, nós também."

means 'to pull of chunks.' Abipón has *l-aoel-ag-Ra* 'POSS.INDEF/3SG-heart/guts-PRODUCT-ABSTR.' Guaraní *ape* colexifies 'back' and *pire* 'movie.' Maxakalí *xax* also means to "seek, hunt, long for." Piro *mta* may also refer to a 'surface' or 'mat,' the Rama term *sarpan* also to 'bast,' Yanomámi *hisi* also to a 'wasp,' and Wayampi *pilɛ* also to any sort of covering of artifacts. Yanomámi colexifies 'bark' with 'bud,' Bwe Karen with "to scratch, claw, maul," Lenakel with 'cover of book' (the term is said to be capable of referring to "any ... outside covering" in the source), and finally, Bislama *skin* is also used with the meanings "very strong, effective, very good, cool."

4. The Bay

Representation: 32%
Motivated: 90.3%
Thereof Analyzable: 47.3% Thereof Colexifying: 44.8%
Thereof by Contiguity: 69.2% Thereof by Similarity: 13.9%
Recurrent associated meanings: corner, sea, harbor, lake, valley, crooked/crookedness, elbow, mouth, breast, inside/interior, bend of river

Five sampled languages -and this already is the most frequently occurring association- colexify 'bay' (or 'inlet,' 'gulf' which were accepted as proxies) with 'corner' more generally. These are Nunggubuyu, Meyah, Khalkha, Toba, and Hawaiian (where the relevant term may also refer to a 'cove,' 'indentation' and 'cell of beehive' inter alia). Tetun has the complex terms *tasi-lidun* 'sea-corner' and *tasi-sikun* 'sea-elbow/corner' (in addition, the language also has the terms *tasi-keta* 'sea-to.separate,' *tasi-nanál* 'sea-tongue' and *tasi-soman* 'sea-companion;' for the second term mentioned, compare colexification of 'elbow' and 'cubit' in Khalkha). For the association with 'corner,' compare that with 'curved shape' reported by Buck (1949: 39) for Indo-European and the somewhat similar analyzable terms *kjamtso-kʰʌ* 'ocean-edge' in Manange, colexification of 'to bend' and 'crooked, winding' in Mandarin, and *helodranomàsina* /hèloka-ranomàsina/ 'crookedness-sea' in Malagasy. In four sampled languages, terms for 'bay' are lexically connected with the meaning 'lake,' by colexification in Nez Perce, Santiago Mexquititlan Otomí and Rotuman (in the latter language by the analyzable term *tạn häe* /tạnu häe/ 'water contain;' the term can also refer to a 'puddle'), while Berik has the complex term *fo boga* /fo bogana/ 'river/lake calm.' Again four sampled languages, Burarra, Lavukaleve, Lesser Antillean Creole French, and Lenakel colexify 'bay' with 'harbor' (compare again Buck 1949: 39 for evidence from Irish and Germanic); in Burarra the relevant term is semianalyzable and contains a verb meaning 'to be helpless.' Another term for 'bay' in which the semantics of the constituents suggest that 'harbor' is a secondary meaning is Miskito *slaup yukuwaik* 'boat shelter.' Lenakel *nouanhulin* is analyzable as /noua-nhul-n/ 'opening-mouth-3SG.POSS.' For this term, compare also Tehuelche *k'on k'ork'en* 'bay, shore of river' which may alongside *k'on ~ k'o:n* 'river' contain *k'onk'en ~ konk'en ~ konken* 'mouth' (or perhaps *k'o:r ~ k'or ~ k'oʔr* 'pasture'?), as well as the common 'mouth'-metaphor in terms for estuary reported in section 20. Two sampled languages of Europe, Basque and Greek, colexify 'breast' and 'bosom' (Basque also 'stomach, guts' and Greek also "apopleptic fit"); this is a pattern peculiar to Europe (see Buck 1949:

38-39 for further Indo-European evidence and § 6.4.3.3. for some details on the history), and in three other sampled languages, Central Yup'ik, Hawaiian, and Takia, associations with 'valley' or 'depression' are found: Takia *ilo-* colexifies "inside, emotions, valley, bay," Hawaiian *Hono-* colexifies 'bay, gulch, valley' (but occurs only as part of place names), and in Central Yup'ik (Yukon dialect), *ilutak* contains the base *ilu-* 'interior' and colexifies 'valley, dip.' Somewhat similarly, *ilutuqaq* /ilutu-tuqaq/ 'be.deep-one.that.has.to.a.large.degree' colexifies 'bay' with 'depression' and 'deep place' (there are also the terms *kangiqaq* /kangiq-qaq/ 'headwaters-area.of.possessor' and *kangir-rluk* 'headwaters-one.that.has.departed.from.its.natural.state;' the full meaning range of *kangir-* is "meaning, principle, source, headwaters of river, beginning" and "strait of water"). Similarly, Rotuman colexifies 'bay' with 'hollow, cavity, recess.' Yir Yoront *larrngarrp* (containing *larr* "place, site, tract, estate, country") and Khalkha colexify 'bay' with 'bend of river.' Upper Chehalis *s-qíwɬ=či* is analyzable as 'CONTINUATIVE-stink=water/river,' and also denotes 'saltwater' and 'sea, ocean;' colexification with 'sea' is also found in Nez Perce.

Other associations include: Yoruba has the (suspiciously long) terms *i-ya-wọ omi òkun sinu ilẹ̀* 'NMLZ-to.branch-to.enter water sea into land' (compare also Japanese *irie* containing *ir-* 'to enter') and *àlafo ti o wà lãrin ọwọ̀ meji* 'space REL 3SG exist between pillar two.' Muna *kolowa* also denotes a 'path in a cave,' and Nunggubuyu *-d̲uṉ-* also a 'burrow,' while Chukchi (*ka) ańee irɣin* contains *irɣin* 'shore.' Kolyma Yukaghir *öge:daŋil'* "semi-circular bay without a estuary"contains *aŋil'* 'estuary.' Welsh *llawryf* also means 'laurel,' and *cilfach* also can refer to a 'nook' or 'recess.' Nuuchahnulth *hita-ċus* is analyzable as 'LOC-dig.' Fijian *toba* contains *tō* 'filled, saturated with water or milk' (*toba* also denotes a plant), and for Great Andamanese *tōngmûgu* 'bay,' compare *tòng* 'leaf' and *mûgu* 'face.' There is a semianalyzable term where the identifiable constituent means 'water, fresh water' in Lenakel. Finally, Rotuman *popo* also means 'rotten, decayed' inter alia.

5. The Beak

Representation: 73.6%
Motivated: 66.8%
Thereof Analyzable: 15.9% Thereof Colexifying: 52.4%
Thereof by Contiguity: 3.2% Thereof by Similarity: 57.5%
Recurrent associated meanings: nose, mouth, snout/muzzle/nozzle, end/point, bird, tooth, lip, prow of canoe, thorn, peck, toe of shoe

To convey the meaning 'beak,' languages overwhelmingly choose body-part metaphors, at least where terms are synchronically motivated. Frequent are associations with the 'mouth.' The association is realized formally mostly by colexification, as found in Buli, Efik, Khoekhoe, Swahili, Baruya, Kyaka (where the relevant term in fact means 'wide beak' specifically), Toaripi, Abzakh Adyghe (among other meanings), Nivkh, Kashaya (the relevant term *haʔbo* is analyzable as /ʔaha-ʔbo/ 'mouth-enlarge/swell.up' and means 'protrusion of the mouth' and 'external mouth'), Yaqui, Hupda (by the term *nɔg'ŏd*, perhaps analyzable as *nɔg-'ŏd* 'mouth.related-inside'), Fijian, Malagasy, Rotuman, Samoan, and Tetun

(colexifying also 'bank, shore;' secondary associations in some languages that are likely genuinely linked with 'mouth' rather than 'beak' are not discussed here unless relevant, see section 124 for these). Six lanuages have morphologically complex terms with 'bird' acting as a contiguity anchor, such as Guaraní *guyra-juru* 'bird-mouth.' These are Kanuri, Mbum, Yoruba, San Mateo del Mar Huave (where 'animal mouth' more specifically is the meaning of the constituent), Guaraní, and Lenakel. Other complex terms where the second element is not 'bird' include Ngambay *ngàw-tà* 'claw-mouth,' Wappo *naphúċe*, analyzable as /nán-phúċe/ 'mouth-poker,' and Kiliwa *ha?=phaqy*, which contains *ha?* 'mouth' and *phaqy* 'pointed' (moreover, Hani *meiqtuv* might be analyzable as *meiq-tuv* 'mouth-peck,' although *meiq* only occurs with the meaning 'mouth' in conjunction with *baoq*, a classifier for hollow things, and both *meiq* and *tuv* have other meanings; a semianalyzable term with 'mouth' is found in Japanese). 'Beak' and 'end' or 'point' are colexifed in Buli, Efik, Ngaanyatjarra (here the term also means 'face' and 'nipple' inter alia), Abzakh Adyghe (also colexifying 'beginning' and other things), Basque (also colexifying 'front, façade' and 'bit smidgeon, a little bit'), Ket (also colexifying 'top' and 'protuberance'), Itzaj, Quileute, Bororo, Guaraní (here also colexifying 'smoke' and 'vapor'), Piro (also colexifying 'arrowhead' specifically), Tehuelche (the relevant term is *ʔor ~ ʔol ~ or*; note the similarity to *ʔor ~ ʔoːr ~ or ~ oːr* 'tooth;' there is another phonologically similar term meaning 'perhaps'), and Yanomámi. There is a semianalyzable term for 'straight beak' with the identifiable constituent meaning 'point' in Great Andamanese (also, in Cubeo, the root *cãpi-* occurring in *cãpibo* 'beak,' when suffixed with a different classifier, assumes the meaning 'point, headland,' which is an association found by colexification also in Fijian). At least for some of these languages, the association might be secondary, since the relevant terms in these languages also colexify 'nose' in addition to 'end,' and this association is in fact the second most frequent lexico-semantic tie to be observed, occurring by colexification in as many as 29 sampled languages (Yoruba, Gurindji, Ngaanyatjarra, Kosarek Yale, Abzakh Adyghe, also colexifying 'beginning,' Chukchi, Khalkha, Kildin Saami, Sora, Chickasaw, where the relevant term may also refer to the 'bill of a hat,' Itzaj, Lake Miwok, Nez Perce, Quileute (where in fact 'nose' and 'mouth' are colexified), Wintu (which also colexifies 'stone, rock' and 'bare protuberance generally'), Copainalá Zoque, Aguaruna, where the relevant term also colexifies 'wax,' Cavineña, Guaraní, Huambisa, Macaguán, Miskito, Piro, Tehuelche, Tsafiki, Wayampi, Fijian, which also colexifies 'cape' and 'mountian peak' inter alia, and Hawaiian. An interesting situation is found in Yir Yoront, where "bottom of bird's beak" is colexified with 'mouth,' and "top of bird's beak" with 'nose.' Five languages, Laz, Cheyenne, Kiowa, Santiago Mexquititlan Otomí, and perhaps Bora have analyzable terms involving a constituent meaning 'nose,' the contiguity anchor being 'bird' in Laz, Cheyenne, and Kiowa, and 'bone' in Santiago Mexquititlan Otomí, while in Bora, the relevant term is perhaps analyzable as *túju-wa* 'nose-CL.PLANK.' Another frequent metaphor-driven pattern of colexifcation is that of 'beak' with 'tooth,' found in Baruya, Kyaka (here, the term denotes a 'narrow beak' specifically and also colexifies 'food' and "sharp, biting, erosive"), Itzaj (by a different term than that mentioned above; this one also colexifies 'grain of corn'), Arabela, Bora, Cashinahua, Lengua, and Rama. Compare also Embera *kidhá* 'tooth' with *kidá* 'beak.' White Hmong has *kaus ncauj* 'beak mouth' for both 'beak' and 'fang, tusk, canine

teeth,' while Highland Chontal has *lihay gaga* /lahay ɬaga/ 'tooth bird' and Chayahuita *inaira natë'* 'bird tooth.' This pattern is obviously very common in languages of South America. An association exclusively occurring by colexification in the sample is that with 'lip,' found in Swahili, Kyaka (colexifying "wide beak" more specifically), Carrier, Central Yup'ik, Jarawara (where the relevant term may refer to 'fruit' and 'lump' inter alia, likely due to the frequent homonymy in the language mentioned by Dixon 2004), Yanomámi (here the colexification is with 'lower lip' specifically), and Takia (here it is with 'pouting lips' more specifically).

A subsidiary pattern is the association with 'snout,' 'muzzle,' or 'nozzle,' found by colexification in Buli, Khoekhoe, Yoruba, Muna, Khalkha (where *xosiɣu(n)* also may refer to 'chatter,' the 'bow of a vessel,' the 'spur of a mountain,' a 'cape, promontory' or 'peak,' and other things), Haida (also colexifying 'mouth of vessel'), Kashaya, Lesser Antillean Creole French, Nez Perce, Quileute, Copainalá Zoque, Guaraní, Huambisa, Jarawara, and Hawaiian (note that in Nez Perce, Quileute, and Hawaiian, 'nose' is also colexified and in Jarawara also 'upper lip' is; in the other languages the association is not at the same time with other pertinent meanings). In San Mateo del Mar Huave, this association is realized formally by a complex term, namely *ombeay quiec* 'animal.mouth bird.' Possible other body-part metaphors are found in Nunggubuyu, where *dhamunun*^{g} is derived from *lhamunun*^{g} 'chin' (which is in turn related to *lha* 'mouth'), and the Sko term *lóeto* resembles *loe* 'ear.'

Metaphorical associations not based on the transfer of a body-part term to the 'beak' are the colexification with 'thorn' in Kyaka (where it is colexified with 'narrow beak' more specifically), Bororo, and Lengua, and that with 'prow (of canoe)' in some South American languages, namely Aguaruna, Bora, Guaraní, Huambisa, Piro, and Wayampi (here by the analyzable term *sĩ-ngɛ* 'nose/beak-old'), as well as Hawaiian. The cluster in South America is suggestive of an areal pattern of South America, and indeed it may be so, but Blust (2009: 314) reports that this association is also common in Austronesian languages, as suggested by its presence in Hawaiian in the present sample. San Lucas Quiaviní Zapotec and Hawaiian colexify 'beak' with 'handle of pitcher' and 'beaker of pitcher' respectively, and in Khalkha and Hawaiian, there is colexification with 'toe of shoe' (the relevant Hawaiian term may also refer to the "thick pearl of pearl-shell shank").

The only recurrent pattern in the sample that is not metaphor-based is the association with verbs meaning 'to peck,' and even this pattern is highly restricted areally, namely to Western North America. In Upper Chehalis, the 'beak' is called *ɬáƙʷmṅ* /ɬáƙʷ=mṅ/ 'peck=INSTR,' in Nuuchahnulth *ƛ̓upkʸak* is analyzable as /ƛ̓upk-ʸakʷ/ 'peck-instrument,' and in Ineseño Chumash, the same root is used for 'beak' and 'to peck' (but note the evidence from Hani discussed above and that Tetun has the redundant complex term *ibun-tutu-n* 'mouth/beak-peck-SINGULATIVE').

Still other associations include: Khoekhoe *ammi* also means 'deep furrow,' Kyaka *nenge* also 'sharp' inter alia, Greek *rámfos* also 'burner, jet (of light),' Welsh *pig* also 'spout,' and Wintu *sO* also 'stone, rock' and *sono* also 'point of rock.' Arabela colexifies 'beak' and 'claw of crab,' and Bororo has one term colexifying 'beak' with 'plait' and 'fibre,' and another one colexifying it with 'sting,' both times among other colexified meanings. Tsafiki uses the same term to refer to the beak of a bird as well as to a 'box,' and Wayampi *sĩ* also

means ‘white.’ In Hawaiian, the semantic range of the relevant term *ihu* also includes “mouth or entrance, as of a harbor, river, or mountain pass or gap” inter alia.

6. The Bird

Representation: 98%
Motivated: 20.9%
Thereof Analyzable: 10.9% Thereof Colexifying: 10.7%
Thereof by Contiguity: 9.2% Thereof by Similarity: 1.6%
Recurrent associated meanings: animal, chicken/fowl/duck/goose, fly, wing, airplane, penis, insect, feather, game, small/little, sky

Terms for ‘bird,’ if they are motivated, are often related to the next higher level in the ethnozoological taxonomy, namely ‘animal.’ In fact, Upper Chehalis, Highland Chontal, Huambisa, Jarawara, as well as Fijian, Rotuman, Kapingamarangi and Samoan, have terms which colexify the two ranks in one term, thus putatively being on the stage of lexical expansion in which terms on the life-form level come to include the unique beginner rank (Samoan also colexifies ‘cattle’ specifically as well as ‘thing’ generally; Highland Chontal also colexifies ‘bug’ and may be an instance of the “wug”-category as defined by Brown 1984; note also that Nunggubuyu, very similarly, has a term also capable of referring to ‘terrestrial game animals’). More frequently, however, are complex terms for ‘bird’ where one of the consituents is ‘animal’ and the other serves to specify and narrow down the semantic range. Terms where the second element is ‘to fly’ are found in Buli (*jayirim* /jaab-yirim/ ‘animal/thing-flying’), Nivkh (*pyi-ṇa* ‘fly-animal’), and San Lucas Quiaviní Zapotec (*ma'anyarzuh* /ma'an-rzuhuh/ ‘animal-flies;’ the language also has the alternative term *ma'any zhaybàà'* ‘animal sky’). However, there are also terms of the derived type, as in Sahu *sosoloro* (*soloro* ‘fly’), Nez Perce *weyíxnikeʻt* /weye-lixnik-eʻt/ ‘fly-move.around-AGT,’ Nuuchahnulth *maamaati* /maa-mat-ǐˀp/ ‘RED-fly-THING....ED,’ and Bora *wahpéé-be* ‘fly-CL.M.S.’ Furthermore, semianalyzable terms where one of the constituents means ‘fly’ are found in Basque, Central Yup’ik, Cavineña, and Toba (here, the indentifiable constituent means more specifically ‘to fly alone’), and in Wintu, the identifiable constituent of an idiolectal word for ‘bird’ is “to glide, spread the feathers to glide” (see also Buck 1949: 183 on this association in Greek and some Celtic and Germanic languages). Frequently, terms involving ‘wing’ are also found (as is also the case in Greek and Sanskrit, see Buck 1949: 183 and § 5.4.2.7.2.). Complex terms with ‘animal’ being the second constituent are Yir Yoront *minh-puth lon* ‘animal-wing with’ and Ket *keŋassel*, analyzable as /kēŋ-assel/ ‘wings-wild.animal.’ Other terms in which the meaning ‘wing’ is used include Buin *perekupa*, containing *kupa* ‘wing’ and presumably *pere* ‘to turn, roll’ (the term also means ‘go hunting for birds’ as a verb), Khalkha *zigyr-ten* ‘wing-COLL,’ and Central Yup’ik *yaqulek* /yaquq-lek/ ‘wing-one.having’ (dialectally meaning ‘angel’). There is a semianalyzable term in Miskito, and the association is diachronically detactable also in Dongolese Nubian. In Hupda, the ‘bird’ is called *hũ-tæ̃ˊh*, perhaps analyzable as ‘animal-small,’ and a term with the same structure is encountered also in Miskito. Complex terms of the derived type in which the meaning

'feathers' figures are found in Carrier, Kiliwa, and Abipón; note also Cahuilla *wíwikmal* ~ *wíkitmal* 'bird' and *wíkił* 'feather' (as well as *málmal* 'butterfly'?).

A relatively frequent association is with 'chicken,' 'fowl,' 'duck,' or 'goose' occurring by colexification in Basque, Chukchi, Greek (see also Buck 1949: 183 for evidence from Ancient Greek), Khalkha, Welsh, Nez Perce, Central Yup'ik, Abipón (also with 'stork' specifically), Lenakel and Tetun. In addition, in Muna *manu* is 'chicken' (compare the apparent cognacy with the 'animal'-term in Rotuman and other Austronesian languages mentioned above) and *manumanu* is 'bird,' while Sahu *namo diwang* is analyzable as /namo diwanga/ 'chicken sky.' This is interesting in the light of Berlin's (1972) report that in Tzeltal the indigenous term for 'bird,' *mut*, experienced a semantic narrowing induced by the introduction of the domesticated chicken by the Spaniards to the effect that its meaning potential was reduced to 'chicken.' In the modern language, 'bird' is expressed by the compound *te?tikil mut* 'forest chicken.' A similar "markedness reversal" (Witkowski and Brown 1983) may have taken place in Sahu and likely also in Muna (but note that there are languages, such as Buli, in which 'fowl' is specifically excluded from the semantic range of the general 'bird'-term).[3] In three sampled languages, Ancash and Imbabura Quechua as well as Tetun, terms for 'bird' are metaphorically extended to 'penis' (a pattern known also in other languages, compare e.g. English *cock*), and in another three, Nunggubuyu, Jarawara, and Lesser Antillean Creole French, terms also mean 'game' ('terrestrial game animals' more precisely in Nunggubuyu). Swahili, Mali, Yir Yoront, and Toba colexify 'bird' with 'airplane' (Mali 'large bird' specifically; on complex terms for 'airplane' on the basis of 'bird,' see section 72). A Khalkha term colexifies 'bird' with 'child,' and the Bwe Karen term *cuɓápho* contains *-pho* 'child.'

It should also be pointed out that in some languages, 'small bird' and 'large bird' are lexically distinguished; this pattern is found in the sample in Rendille (where the term for 'large bird' also denotes the 'vulture' specifically; compare Sasse 2002: 1053 on this in Cushitic lanugages in general), Carib, and Yanomámi. In Buli, *jayirim* may also refer to an 'insect,' and the same is true of the Miskito term *daiwan lupia* 'animal small' and the Rotuman term *mạnmạnu ferfere* 'bird/animal restless.' Hani colexifies 'bird' with 'meat,' for which compare the likely relationship between Yei *yarmakər* 'bird' and *makər* 'flesh.'

Further unique associations are rare: In Buli, *nuim* is also the name of a mark made in the skin believed to act as a remedy against a disease caused by a particular bird. Efik *i'nuën* can also refer to a kind of magical conjuration. In Khoekhoe, *anib ~ anis* 'bird, cock' is derived from *ani-* 'decorate, adorn, color in, dress up,' Ngambay *yèel* also means 'to feel sick,' while in Kwoma, the relevant term also includes 'flying foxes' in its denotational range and that in Kyaka also 'bat' alongside other meanings (but not flying insects). The relevant Muna term also denotes a children's game involving flying objects, and that in Toaripi also a kind of flying fish. Sko *táng* also means 'fishing net,' while Basque *txori* also means 'thief,' 'mumps,' 'hinge,' 'eyelet,' 'dickey' and 'bun.' Itzaj *ch'iich* also means "pick out (small things), delouse," and for Kiowa *kuɑtou* compare *tou* 'house.' Wintu *ċil* also means

[3] The precise denotatum of English *chicken* is subject to some dialectal variation and there thus may be ambiguities here due to the use of English as the metalanguage.

'eclipse' and 'bear.' The quite obviously onomatopoetic Wintu term *tili·tilit* also denotes the 'snipe' specifically, while Yuki *či·mit* also the 'blackbird' specifically. Arabela *shiyojua* contains the classifier *-jua* which is glossed as 'tierra, volador' in the source, and for Maxakalí *putuxnãg* compare *putux* 'heavy.' Sáliba *llĩde* also denotes the 'moriche palm,' while Wayampi *wɨla* also denotes a round dance consecrated to birds, as well as, presumably by homonymy, 'tree, wood.' Yanomámi *yõõ pesi* 'big bird' contains *pesi* 'sheath, nest' (the term also has other meanings), Bislama *pijin* is rarely used for Bislama, but commonly employed to refer to "Melanesian Pidgin as spoken in Solomon Islands and Papua New Guinea," i.e. probably Tok Pisin. Hawaiian *manu* also denotes "any winged creature" as well as 'wing of a kite,' Lenakel *menuk* is also the name of a child's illness said to be caused by birds. Manange *1ŋima* also means 'gill, ear,' Sedang *chêm* is also the Sedang name of the star Antares, and Yay has phonologically identical terms for 'bird' and 'to hoe.'

7. The Blossom

Representation: 93%
Motivated: 23.4%
Thereof Analyzable 7.7% Thereof Colexifying: 15.7%
Thereof by Contiguity: 4.2% Thereof by Similarity: 7.3%
Recurrent associated meanings: plant, bud, fruit, eye, pretty, shine, burst, pod, child, lid/cover, sprout, rose, picture, grass, head, yellow

Lexico-semantic associations of the meaning 'blossom' ('bloom,' 'flower') are manifold, though none of them is particularly frequent cross-linguistically. Six languages, Efik, Khoekhoe, Yoruba, Baruya, Itzaj, and Pawnee colexify 'bud' by temporal or provenience contiguity (and Pawnee in addition 'flower bud' specifically). Another Itzaj term colexifies 'flower' and 'sprout, seedpod,' and 'pod' and 'flower' are also colexified in Rao. Efik (by a term derived from a verb meaning 'to peel' and colexifying 'seed' inter alia), Kaluli, Kyaka, Sahu, and Samoan colexify 'flower' with 'fruit' ("inedible tree fruits" in Kaluli specifically; Samoan also colexifies 'egg' and 'products,' compare the colexification with 'roe of fish' in Rotuman). Metaphor-driven conceptualizations, however, are more frequent than contiguity-based ones. Badaga and Hawaiian colexify 'flower' and 'child(ren)' (the Badaga term also means 'shoot' and 'harvest,' and the semantic range of the relevant Hawaiian term also includes "tassel and stem of sugar cane" and 'arrow, dart' inter alia), Buli and Bororo colexify 'flower' and 'stomach' (the Bororo term also means 'liquid, blood' and 'shininess' inter alia, and the Buli term also includes 'pregnancy' alongside other meanings in its semantic range, see discussion of connections between 'stomach' and 'womb' in section 148). In Pawnee, the generic term for 'flower,' *kiriktarahkataaru?,* which also, however, denotes the 'sunflower' more specifically, contains *kirik-* 'eye' and *rahkataar* 'be yellow' ('flower' is colexified with 'yellow' in Malagasy). Furthermore, Wintu *tuh* 'flower, bloom, blossom' betrays a lexical connection to *tuh* 'eye.' The lexicographer (Pitkin 1984) remarks: "parallel to English daisy, i.e., day's eye." Similarly, in Central Yup'ik, the base *uite-* in means 'to open one's eyes, to wake up, to bloom (of flowers),' and parallely, Burarra

colexifies 'to have eyes open' with 'be in bloom,' by the analyzable term *ga-na ~ ja-na* 'take-see/look.at.'

In Yir Yoront, *par-thaw* is analyzable as 'head/top-mouth,' and in Cheyenne, *tsé-pêhévêstséavó'o'e* also contains *htséá* 'head;' lexicographers state that a literal morpheme-by-morpheme translation of the entire term would be "that.which.is-pretty-headed-growth" (cf. Central Yup'ik *nau-cetaaq* 'grow-something.used.to.cause.to'), and also note that "[it] is quite likely that there was no historical word for 'flower'" (for 'pretty' as the meaning of a constituent, compare colexification of 'pretty' and 'flower' in Miskito, Ket *aqta daan* is 'pretty/good grass,' and for the association with 'mouth' compare Blackfoot *waapistsisskitsii* 'flower' with *waapitstsi* 'empty one's own mouth of'). The comparison is likely based in both cases on the fact that both 'head' and 'flower' are roundish structures that are found on top of the human body or a twig of a tree or bush or flower respectively. This is further corroborated by Rotokas *kokookua* 'open flower,' which contains *kokoo* "food on plate, something which is placed on top of something else" (see Buck 1949: 525 for a similar association with 'point' and 'edge' in Indo-European). Rama and Yay colexify 'flower' with 'lid' or more generally 'cover' (the Yay term colexifies in addition 'wall of a house'). In Dongolese Nubian, *ǩǧ* also means 'fire, flame' and 'light,' and perhaps similarly, Yoruba *itanna ewéko* is analyzable as 'lamp.lightning plant,' Muna *kambea* is derived from *mbea* "glow, shimmer, shine, twinkle," and in Lake Miwok, 'to bloom' is colexified with 'to shine' as of the sun inter alia (see Buck 1949: 526 for Indo-European). Furthermore, for reasons unknown, two languages of Eurasia, Khalkha and Kolyma Yukaghir, colexify 'flower' with 'picture;' in Khalkha, the relevant term also denotes 'smallpox,' and another Khalkha term denotes in addition the concepts 'comb of cock' as well as 'club in cards,' in Kolyma Yukaghir the relevant terms also means 'embroidery,' 'ornament' and 'color.'

Itzaj and Lake Miwok colexify 'flower' with 'sprout.' Comanche *mubu̶tsaakatu̲* contains a verb referring to something blowing up or bursting, and 'to bloom, blossom' and 'to burst' are colexified directly in Nunggubuyu and Itzaj. The Dongolese Nubian and Abipón terms also denote the 'rose' specifically (according to Wehr 1976: 1061 this is a pattern of autohyponymy also found in Standard Arabic, so that areal spread or even direct borrowing into Dongolese Nubian is at least a possibility).

Before listing patterns found only in one sampled language each, it must be pointed out (as already done in chapter 3) that a problem with the concept 'flower' is that the English word *flower* is ambiguous between the reading 'blossom' and 'smallish blossoming plant' (German *blüte* vs. *blume*). This hampers systematic analysis, as lexicographers often do not make clear which of the two meanings of English *flower* is meant. It cannot be excluded that colexification with 'plant,' found in Meyah, Badaga, Kildin Saami (here also with 'grass' and 'vegetable'), San Mateo del Mar Huave, Santiago Mexquititlan Otomí, and Central Yup'ik is in some cases due to lexicographers selecting the *blume*-sense of 'flower,' and the same is probably true for terms such as Ket *aqta daan* 'pretty/good grass.' Another point worth mentioning is that terms for 'flower' are in many languages deverbal in nature, with a verb meaning 'to bloom, blossom' being derivationally basic (see again Buck 1949: 526 for Indo-European).

Other associations include: Hausa *huda*, a term of the Kuda dialect for the flower of any plant other than that of the tobacco plant, also means "be slightly split," "ooze with water," and 'to pierce.' Ngambay *pútu* also means 'white hair.' Swahili *ua* also means 'fence' and 'yard, court,' while Yoruba *òdòdó*, as an adjective, denotes the color 'scarlet.' Kwoma *she* also means 'ash' and 'faeces,' and Kyaka *jingi* also 'waterfall, rapids' (for the meaning 'flower,' *isa* 'tree, bush' may be used in apposition). Nunggubuyu *-dhabag* is derived from *lhabag* 'feather;' on the basis of this term, there is the complex verb *-lhababida-* meaning 'to be in bloom,' presumably originally containing *-wʔida-* 'to be new moon' (the lexicographer notes that "the semantics make a synchronic segmentation questionable"). Similarly, Buin *paru* 'feather, plumage' is also glossed as "(used for) flower." Sko *pang* in *rípang* 'flower' (where *rí* means 'tree, wood, scales of fish') is identical segmentally to *pang* 'pus.' Cheyenne *véhpôtse* "primarily means 'leaf' but is sometimes extended to mean 'flower'" (compare also Nivkh *eŋv tjomr* 'flowering leaf'). Muna *wuna*, a term for 'bloom' which is however not used for the flowers of most trees and flowers, also denotes "[t]he island of Muna," as "according to legend it is called Muna (=flower) because the first inhabitants saw a stone with flowers growing on it." Badaga *hu: ~ u:* also means 'tail-side of coin,' and Japanese *hana* also 'nose, nasal mucus' (the terms are prosodically different though, and in the 'flower'-sense, *hana* may be an early borrowing from Austronesian). The literal meaning of *héováéstse tséhetó'o'ee'êstse*, a Cheyenne term for 'flower,' is stated to be "various kinds how they grow." Chickasaw *pakạli'* is derived from *pakạli-* 'to open, bloom,' and similarly, Lake Miwok *táke ~ take*, a verb meaning 'to bloom,' also may mean 'to come out, emerge' inter alia. Itzaj *wach'äl* is derived from *wach'-* 'to loosen, disentangle,' and Wintu *lul* is connected lexically to *lul* 'long cylinder.' San Lucas Quiaviní Zapotec *gyìà'* can also refer to a mixture of corn meal floating on top of tejate (a traditional Oaxaca beverage made from maize and cacao). Aguaruna *yagkúg* is also a man's name (compare also the segmental similarity to *yagkú* 'yellow'), Cashinahua *jua* also denotes 'corn silk,' and Cayapa *llullu* appears to be a reduplication of *llu* 'penis, wood.' Chayahuita *nitërinso'* is derived from *nitërin* 'bear fruit,' and Jarawara *mowe* also denotes the 'Brazil nut.' Tsafiki *lulí* may also refer to a 'cockscomb' and is related by unknown means to *luban* 'red,' while Toba *lasoviaxa* is a nominalization of *asovi*, colexifying 'to bloom' with 'to fray, frazzle.' Yanomámi *horehore* appears to be reduplicated from *hore* 'hidden below.' Hani *yeiv* also means 'not firm,' Tetun *funan* can also assume the meaning 'mildew' and 'product, interest,' and *aifunan*, a compound with *ai* 'tree, plant,' can also refer to one's 'girlfriend.' Vietnamese *bông* also denotes 'cotton,' Yay colexifies 'flower' with 'bone,' Bislama *flaoa* also denotes "any plant grown for decoration (as against plant grown for food)" and Lesser Antillean Creole French *fle* is also used with the meanings 'best' and 'choice.'

8. The Branch
Representation: 95%
Motivated: 40%
Thereof Analyzable: 12.5% Thereof Colexifying: 27.5%
Thereof by Contiguity: 9.9% Thereof by Similarity: 28.3%

Recurrent associated meanings: arm, hand, leg, wing, tributary, horn, knot in tree, tree, fork/divide/separate/branch off, bush, shoot/sprout, point/end, fin, wing of army, descendants, prong, leaf, member, root, bone

The most frequent lexico-semantic association for 'branch' is metaphorical in nature and consists of lexical ties with either 'arm,' 'hand,' or both (assuming in the following discussion that terms glossed as 'limb' do not exhibit any of the patterns of colexification). This pattern is also attested diachronically in the history of Germanic according to Buck (1949: 523). Formally, both colexification and morphologically analyzable terms are amply attested. All in all, the association with 'arm' is found in 29 sampled languages, by colexification in Koyraboro Senni, Buin, Burarra, Kwoma, Mali, Ngaanyatjarra, Khalkha (more generally, 'limb' is colexified here), Ineseño Chumash, Itzaj, Kashaya, Kiliwa (by the analyzable term *t+haq=tay* 'OBJ+bone=be.large;' 'branch' and 'bone' are colexified, inter alia, in Bororo), Quileute, San Lucas Quiaviní Zapotec (colexifying 'upper arm' and 'upper arm of front leg of animals' specifically), Carib, Cashinahua, Jarawara, Bislama (among other colexified meanings), Great Andamanese, Fijian (also colexifying 'upper arm' specifically'), Hani (also colexifiying 'branch of clan'), and Lenakel, and by analyzable terms in which 'tree' acts as a contiguity anchor, as in Yir Yoront *yo-puth* 'tree-arm,' in Anggor, Dadibi, Yir Yoront, Highland Chontal, Santiago Mexquititlan Otomí, Chayahuita, Kaingang, and Yanomámi; furthermore, a semianalyzable term where one of the constituents is 'arm' is found in Yana. With 'hand,' colexification is found in Koyraboro Senni, Yoruba, Buin, Kwoma (also with 'handprint' and "artist's personal style of painting and carving"), Mali, Ineseño Chumash, Itzaj, Quileute, San Lucas Quiaviní Zapotec, Jarawara, Bislama (again, among other colexified meanings), and Lenakel, and analyzable terms are present in Mbum (*ndòk-kpù* 'hand-tree'), Anggor, and Kosarek Yale. Moreover, there is a semianalyzable term in Pipil. Now, languages frequently colexify 'hand' and 'arm' (Brown 2005b), and where the above groups overlap, this is due to such colexification.

Other complex terms with 'tree' as contiguity anchor include One *aila plana* 'tree two,' San Mateo del Mar Huave *omal xiül* 'point tree' (for which compare colexification of 'branch' with 'pointed, pointed object' in Abzakh Adyghe and with 'pointed, sharp' inter alia in Muna), and Tetun *ai-sorun* 'tree-side.' Furthermore, Nuuchahnulth *-maptmapt* is reduplicated from *-mapt* 'plant, bush, tree,' and semianalyzable terms in which one constituent is identifiable as 'tree' (and sometimes 'wood,' see section 65.) are found in Berik, Sko, Waris, Rama, and Bwe Karen. However, 'tree' cross-linguistically not only serves as a contiguity anchor in complex terms, but colexification of 'tree' and 'branch' by spatial contiguity, or more precisely meronymy, is also attested, namely in Ngambay (where many other tree-related meanings are colexified as well, among them 'forest'), Buin (colexifying 'small tree' and a particular tree species more precisely), Welsh, Blackfoot, and Kiliwa. Similarly, in Efik, *okpüt* also colexifies "branches of the young trees lopped off in clearing a plantation" as well as "the trees themselves being left as yam sticks." Moreover, Dongolese Nubian, Muna, and Khalkha colexify 'branch' with 'shoot, sprout,' and Haida *hlq'a7ii* also means 'bush,' "stem (raceme or panicle) of berries or of blossoms grow-

ing on bushes" and 'supporting frame' inter alia; the association with 'bush' is also found in Middle Eastern and Southeastern Tasmanian and Nuuchahnulth.

Interestingly, an alternative to 'hand' and/or 'arm' is colexification with 'leg,' as found in Buli, Khalkha ('limb' generally is colexified here), Arabela, and White Hmong (in Buli also with 'hind leg,' 'foot' and other meanings specifically, and in Arabela also with 'handle of pot,' 'backrest of hammock,' 'grip of bag,' and other things), and with 'foreleg' specifically in Buin, Ngaanyatjarra, San Lucas Quiaviní Zapotec, and Fijian (in Buin also with 'pectoral fin;' 'fin' is colexified with 'branch' inter alia also in Hawaiian). As in Buli, Nunggubuyu colexifies 'branch' with 'hind leg, upper leg' specifically, and has the redundant complex term *-nᵍu-dhalbar* '-guts-upper.leg/hind.leg,' for which compare Maxakalí *mĩm-mãg* 'wood-small.intestine.' Similarly, colexification of 'branch' and 'leg of insect' is found in Kwoma, while Wappo colexifies 'branch' with "many-branched, many-legged."

Other notable metaphor-driven patterns of colexification are that with 'horn' or 'antlers,' occurring in Hausa, Ngambay (colexifying also other meanings), Basque, Kildin Saami (more specifically, "branching at the antlers"), Abipón (here, 'jump' is colexified in addition, and there is a redundant complex term involving 'wood'), and Toba, that with 'wing' in Dongolese Nubian, Buin, Burarra, Kwoma, Ngaanyatjarra, Cashinahua, Jarawara, and Fijian (Dongolese Nubian, notably, is the only of these languages in which the association with 'wing' is surely genuine, that is, in which the relevant terms do not also colexify 'arm'), and that with 'tributary' of a river in Khoekhoe, Baruya, Basque, Khalkha, Arabela, Jarawara, and Toba (and in Burarra, *mernda* ~ *marnda* and *gurnjirra* also denote the 'creek' itself, an association that is at least etymologically also detectable in Nunggubuyu and is also attested in Germanic according to Buck 1949: 524). Semantic extension to 'wing of an army' is documented for Hausa and Hawaiian (and to 'division' in Lesser Antillean Creole French), to 'descendants' in Hausa and Khoekhoe (in which latter the relevant term also can refer to a 'tiller,' 'sucker' and 'water-shoot' and "leg of skin/hide"), to 'member,' as of e.g. a society, in Khalkha and Hawaiian (in both languages alongside other meanings), and to 'prong' in Khalkha and Rotuman (in Khalkha also to 'component' generally, as well as 'department,' 'detachment').

Terms for 'branch' are also frequently associated lexically with meanings such as 'to fork,' 'to divide,' 'to separate,' or 'to branch off' (see Buck 1949: 523). Kaluli *i gasa* is analyzable as 'tree things.that.separate' (the language has another term, *i elé*, where *elé* is identical segmentally to the word for the 'moon'), Kyaka *paka pingi* as 'branched/divided root/rudiment/base,' itself denoting a 'division' or 'fork' as well, and Pawnee *rakitwaiʔuʔ* as /rak-kita-wi-uʔ/ 'tree/wood-be.branching-??-NOM.' Khalkha has a derived term, and such associations are also found by colexification in Badaga, Kildin Saami, Rotuman, and Sedang (furthermore, Sahu *sasalanga* colexifies 'fork in tree, branch, fork in road,' and Muna *tuna* also means 'fork of fingers and toes' as well as 'to bud'). Furthermore, five languages of Western North America, Carrier, Upper Chehalis, Lake Miwok, Nez Perce, as well as the Nunivak Island dialect of Central Yup'ik colexify 'branch' and 'knot in tree' or 'knot in wood,' and two sampled languages, Gurindji and Nuuchahnulth, colexify 'branch' and 'leaf' (Gurindji more specifically 'leafy branch' while Nuuchahnulth also uses the relevant term to denote the 'Kinnikinnick Berry'). Ngaanyatjarra colexifies 'side root,' and Buli 'root'

generally (compare the association with 'root' in Indo-European mentioned by Buck 1949: 523).

Other patterns found in the data are: Efik colexifies 'branch' with 'root shoots of cereals,' and another Efik term, *ñ'kök*, is derived from *kök*, meaning 'to overlay, overlap' and 'to heap up' inter alia. Khoekhoe *ǁnâub* also denotes the 'brim of a hat,' Rendille colexifies 'branch' with 'notch' as well as '(male) dog,' and Yoruba with 'spray.' Muna *karagha* also denotes a "large natural drain, ravine" inter alia, *tuna* also means 'bud' as well as 'to bud, sprout,' and *ragha* also 'plot, field, piece of land.' Ngaanyatjarra *mirna* may also refer to a "pannikin, mug, cup," 'sleeve of dress,' and "steering arm or wishbone of vehicle." Sahu colexifies 'branch' with "midrib of a palm frond," while Abzakh Adyghe *qʷʔətame*, a formally redundant compound of *qʷʔə* 'branch' and *tame* 'wing,' may also refer to 'grape,' and *qʷʔəre* also means 'hard, rough, raspy.' Badaga *moḷe* also means 'nail' and 'peg' and *sui* also 'whorl of hair.' Basque *adar* also can refer to a 'bedpost' or 'chairpost,' as well as a 'lineage.' Khalkha *nailzaɣur ~ nailzur* is analyzable as /nailza-ɣur/ 'swing-INSTR' (the variant *nailzaɣur* also means 'glanders') and another Khalkha term colexifies 'branch' with 'tendril,' while Welsh *cainc* also means 'tune, song' and 'knot.' Nuuchahnulth colexifies 'branch' with 'rind, bark' as well as, by another term, with 'plant.' Wintu *lOb* is said to be connected to *lEw* 'oscillate' as well as to meanings like 'hanging, overhang.' The base *ava-* in Central Yup'ik *avayaq* occurs in other terms with the meaning 'son.' Aguaruna *tsajám(pe)* also means 'finger,' Cashinahua *punyan* also 'sleeve,' Embera *hʉ̶wáte* also 'stick pole' (the meanings are associated with different genders), and Jarawara *mani/mano*, colexifying already 'hand,' 'arm,' 'wing' and 'tributary,' furthermore may refer to a 'watch band' and 'bunch.' Macaguán *pebukrán* might be semianalyzable (*pebʉ́*, 'fruit'), and Sáliba *anojahñojo* appears to contain *anojaha* 'shoulder.' Wayampi colexifies 'branch' with 'head' and other meanings, and Great Andamanese *âkàchâti* appears to be derived by prefixation of a possessive prefix from *châti,* a term for a 'yam species.' Bwe Karen colexifies 'branch' with 'narrowest part of something,' and Fijian inter alia with 'store of the house' and 'page of book.' Kapingamarangi *manga* also means 'to turn brown from heating' and 'to tan,' Hawaiian *lālā* colexifies 'coconut frond' and 'to diverge,' among many other meanings, and *mana* also means 'mana,' 'power,' and 'variant, version,' again among other meanings. Finally, Kapingamarangi colexifies 'branch' with 'branch of government.'

9. The Bud

Representation: 51%
Motivated: 53.5%
Thereof Analyzable: 19.7% Thereof Colexifying: 33.8%
Thereof by Contiguity: 10.5% Thereof by Similarity: 7.9%
Recurrent associated meanings: shoot/sprout, flower/blossom, to sprout, seed, eye, young leaf, child, tobacco, pimple, button, fruit, foam, egg, burst open

Terms for the 'bud' are sometimes associated lexically with the meaning 'to sprout.' Yoruba has *ẽhu ohun ọ̀gbìn* 'sprout thing plant,' Berik *fas ferwesini* /fas ferwesili/ 'zero/NEG sprout,' in Chayahuita *pichopitërinso'* is lexically related to *pichopitërin* 'to sprout,' Tetun

has *ai-tubuk* 'plant-to.sprout,' while 'bud' and 'to sprout' are colexified directly in Hausa and Muna. Colexification with 'shoot, sprout' (and sometimes also 'seedling') is attested in Kyaka, Ngaanyatjarra (also with "small plants emerging from the soil" as well as "budding grain, unripe grain" specifically), Rao, Badaga, Basque, Welsh, Pipil (the relevant term *-mulinka* is derived from *-mulu:ni* 'to swell up, dry out'), Wappo, Lesser Antillean Creole French, Aguaruna, Arabela, Bora, Embera, Guaraní, Ancash Quechua, Yanomámi (which also colexifies 'new branch'), and Tetun, in which latter some of the relevant terms are analyzable: *ai-bo'as* 'plant-burst' and *ai-dikin* 'plant-shoot.' Ten languages in the sample showcase associations between the meanings 'bud' and 'flower' or 'blossom.' These are Efik, Khoekhoe, Yoruba, Baruya, Itzaj, and Pawnee by colexification (in Khoekhoe and Pawnee, relevant terms are deverbal derivatives from verbs meaning 'to bloom,' while in Itzaj, *wach'äl*, which can also mean 'tassel,' seems to be derived from *wach'* 'to loosen, disentangle;' moreover, the Samoan term may optionally be enhanced by an element meaning 'to open'). As for complex terms, the Basque term is metaphorical and based on 'eye' (*lore-begi* 'flower-eye,' note also direct colexification of 'eye' and 'bud' alongside other meanings in Ancash Quechua and Hawaiian as well as Piro *yhale-xi* 'eye-fruit/seed/DIM' which, in addition, colexifies 'ankle' and 'belt buckle'), while Kyaka *jingi ene* is analyzable as 'flower unfinished,' Wayampi *pɔtɨ-yaʔɨ* as 'flower-child' and Kapingamarangi *akai modo* as 'flower unripe.' Furthermore, three sampled languages, Sko, Jarawara (by the analyzable term *aka-bori* 'wear-container/nest'), and Hawaiian, colexify 'bud' and 'young leaf,' and in two sampled languages, Buli and Tetun, a lexical association with the meaning 'burst' or 'burst open' is found - in Buli by colexification (also with 'to pierce,' 'to rise' inter alia), and in Tetun by the complex term *ai-bo'as* 'plant-burst;' furthermore, Piro *muhip-werekatu* also appears to contain *hipwere* 'to burst open.' Colexification with 'fruit' is found in Efik (by a term derived from a verb meaning 'to peel' inter alia) and Sáliba (also with 'seed' in this case, as in Efik and Lesser Antillean Creole French, which latter also colexifies 'germ' and 'rose-bud;' note also Oneida *yonʌ́huteʔ*, containing *-nʌh-* 'seed, grain, oats' and *–N-ut-* 'be attached,' and the Piro term already mentioned above). Efik and Toaripi colexify 'bud' with 'tobacco' ("trade tobacco, tobacco twist" specifically in Toaripi), and Khoekhoe and Ancash Quechua with 'foam' (the Khoekhoe term denotes the bud of *Acacia Watkins* specifically).

Further metaphorical associations include that with 'child,' occurring by colexification in Tuscarora (here also with 'embryo') and Hawaiian (where the relevant term also has a range of other meanings) and by a morphologically complex term (*pɔtɨ-yaʔɨ* 'flower-child') in Wayampi (note also colexification of 'bud' with 'descendant' in Wichí), as well as the colexification of 'bud' and 'pimple' in Fijian and Lesser Antillean Creole French (also with 'protuberance' generally inter alia in Fijian). In two languages of the sample, Lesser Antillean Creole French and San Lucas Quiaviní Zapotec, the relevant terms also mean 'button;' this is likely due to copying of the pattern of colexification in the lexifier language French for the former and to that in the donor language Spanish (*btoony* is a loanword from Spanish *botón*) in the latter language. Finally, Kaingang *jumĩ kỹ jẽ* 'flower bud' looks as if it contains *jẽ* 'tooth,' and the same is true of Meyah *ofoncowú* 'to bud' (*ofon* 'tooth').

Other associations include: Efik *i'köñ* has a wide range of meaning, including 'grass, herbs,' 'vegetables of all kinds,' 'leaf,' and 'tobacco.' Hausa *dum'baru* also means 'lips' as well as "first coming-through of a tooth," while *toho* also denotes a particular drum. Noni *com* 'to bud' also means 'to color,' Berik *koksa* also 'to grow,' Kyaka *kene* may also refer to a "small edible crab, found in some streams," while the Muna term *tuna* may also refer to a 'twig' or 'branch' as well as to the 'fork of fingers and toes.' Muna *lasi* also means 'to weld' (this sense is due to borrowing from Indonesian/Dutch). Ngaanyatjarra *yurnturntu* is also used with the more specific meaning "budding grain, unripe grain." Rotokas *koko ruu* is a complex term consisting of the constituents *koko* 'pour out, spill' as well as 'dish out, portion out' and *ruu* 'enclose,' a variant of which is *kokooko ruu*, where *kokoo* means "food on plate, something which is placed on top of something else." Toaripi *kuku* assumes the meaning 'unopened' when occurring in compounds, Sahu *boro* seems to be derived from *woro* 'wide, spacious,' and the primary meaning of Yir Yoront *nhapn* is 'egg' (note also that Great Andamanese *ârmōl* appears to be derived from *mōl-o* 'egg,' and that a diachronic association between 'bud' and 'egg' is likely in Wintu). Badaga colexifies 'bud' with 'hair knot, braid,' and Basque *ernamuin* also means 'outbreak.' The meaning of Khalkha's *comurliγ* ranges over "bundle, bunch, bouquet" as well as 'anthology,' whereas *nakija* is 'shoot of grass, leaf bud' (compare *naki-* 'to bend; to cave in, be shaky'?). Upper Chehalis *məyspáqnṅ* contains *máy* "begin to, just now, recently, just a little while ago, hardly" and *páqin* 'to bloom.' Arabela *jiyoto* colexifies 'bud' and 'nut,' and Embera *dodháa* also means 'cocoon.' Miskito *kiama namika uruwan ba* is not fully analyzable, but the element *kiama* can also refer to the 'ear,' while Ancash Quechua colexifies 'bud' with 'yerbasanta,' a type of tree. Toba *lỹaxataxa* is derived from a verb meaning 'to raise, grow,' Fijian has *kovukovu*, with *kovu* 'to tie up' being the reduplication base. The term is also used with reference to an enclosed portion of land. Hani colexifies 'to bud' with 'to weed with hoe.' The semantic range of Hawaiian *liko* also includes 'shining, sparkling' and 'fat' inter alia, while *'ōpu'u* is derived from *pu'u* with the basic meaning 'protuberance.' Samoan *moemoe* is a reduplication of *moe*, meaning 'sleep,' and Sedang *rơmôe lóang* also may refer to the "soft, tender part of a plant." Finally, Tetun *tubun*, adjectivally, also means "young, fledgling."

10. The Cave

Representation: 84%
Motivated: 53.3%
Thereof Analyzable: 24.3% Thereof Colexifying: 29.2%
Thereof by Contiguity: 37.9% Thereof by Similarity: 4.5%
Recurrent associated meanings: hole, stone/rock, den/lair, earth/ground, tunnel, valley, pit, opening, house/dwelling, ledge/overhang, in/inside, hill, mouth

As for the 'cave' ('cavern,' 'grotto'), there is one motivational pattern that is by far the most frequent and outnumbers all others in terms of the number of languages it occurs in, namely complex terms of the lexical type where one of the constituents means 'rock' or 'stone' and the other 'hole.' Structurally, these terms obviously vary from language to language. For instance, Hupda has the compound *pǎç məy* 'stone/hill hole/house,' Yoruba

has the phrasal term *ihò inu àpáta* 'hole in rock,' Cubeo *cũ̱racobe* consists of *cũ̱ra* 'stone' and *-cobe*, a classifier for hole-like objects, and in Oneida, *yotstʌhlaká·lute?* consists of the verb *-kahlut-* 'be an opening, be a hole' which has incorporated the root *-itstʌhl-* 'rock, mountain, outcropping' (there is an alternative term where the incorporate is *-uhwʌtsy-* 'land, earth, ground') and accompanying grammatical material, but the source concepts are remarkably constant cross-linguistically. All in all, this pattern is found in 20 sampled languages all over the world: Yoruba, Berik, Toaripi, Basque, Highland Chontal (with some phonological deviations), Haida, San Mateo del Mar Huave, Oneida, Bororo (where *ia* in *ia-ri* also means 'mouth' and 'opening' more generally, compare also Baruya *simaanga* /sɨla-maanga/ 'stone-mouth' and the semantic extension of 'mouth' to 'opening' in general discussed in section 124), Cubeo, Guaraní, Hupda, Rama, Wayampi, Hani (where the association is also realized by means of a classifier), Lenakel, White Hmong, Tetun, and Bislama. In addition, San Mateo del Mar Huave and Tuscarora have complex terms based on roots meaning 'hole' and 'hill,' while Cheyenne *tsévé'evótoo'e*, literally, according to the source "that which is a concave hole," can be extended to mean 'vagina' or 'anus.' A derived term featuring a lexical constituent meaning 'to make a hole in the ground' is featured in Great Andamanese, a semianalyzable term one of the constituents of which is 'hole' is found in Rama, and one with 'stone' is featured in Kemtuik. Alternative complex terms based on 'stone, rock' are Itzaj *'aktun*, perhaps /'aak-tun/ 'turtle-rock' and Chayahuita *na'pitëana,* containing *na'pi* 'rock' and *-ana*, a classifier conveying the notion 'around' (there is also the term *cari-ana* 'precipice-CLASS.AROUND'). In Malagasy, there is a complex term *lavabàto*, consisting of *vàto* 'stone' and *làvaka*, which in fact colexifies 'hole,' 'pit,' and 'cave,' and it is not the only language where the association between 'cave' and 'hole' or 'hollow' is formally by colexification: Efik, Sko, Badaga, Khalkha (colexifying also 'empty'), Kildin Saami, Sora, Acoma (by the analyzable term *pâaniizeeṣa* containing *pâani* 'bag' and *zeeṣa* 'place where'), Cahuilla, Upper Chehalis (where *ʔacƛ̓wál[a]xʷɬ* contains the root *ƛ̓ʷól̓xʷ-* 'hollow' and a marker for stative aspect), Cheyenne, Comanche, San Mateo del Mar Huave, Itzaj, Lesser Antillean Creole French, Quileute, Xicotepec de Juárez Totonac, Wappo, Wintu, Aguaruna, Cavineña, Cayapa, Huambisa, Ancash Quechua, Toba, Yanomámi, Fijian, Hawaiian, Kapingamarangi, Mandarin, and Takia all feature terms that are likewise of this type (Toba and Yanomámi also colexify 'opening' generally). In analyzable terms, an alternative pattern is for the constituent other than 'hole' to mean 'earth,' 'ground' or 'land' rather than 'stone,' as in Cayapa *tu-juru* 'earth-hole.' Such terms are also found in Kaluli, Chickasaw, Oneida, Pawnee (where the relevant term also colexifies 'cellar' or 'storm cellar'), Tuscarora, Bora, and Guaraní (where an additional element meaning 'big' is present), and a semianalyzable term where one of the consituents is 'earth' is also found in Huambisa. A semantic association loosely related to the complex revolving around the meanings 'stone' and 'rock' in some languages is that with 'house' or 'dwelling:' Kyaka has *kana anda* 'stone/rock house' (with both constituents also having other meanings), Sahu *ma'di sasa'du* 'cavern' is analyzable as 'stone/rock ceremonial.house' (more precisely, *sasa'du* is glossed as "ceremonial house with octagonal roof present in each original Sahu village"), and, as already seen, in Hupda *mɔy* in *pǎç mɔy* colexifies 'hole' and 'house.' The association with 'house' or 'dwelling' is found by colexification in Kwoma, while Comanche colexifies 'hole,'

'cave' and 'room.' Burarra (where the relevant term *gun-ngarnama* is analyzable as 'CLASS.DOMESTIC-inner.thigh') colexifies 'room' and 'wall.'

There are also other unrelated lexico-semantic associations for terms meaning 'cave.' Sora *'luŋərən* colexifies 'inside,' while Chickasaw *yaaknanonka'* is analyzable as /yaakni'-anonka'/ 'earth-inside,' Bwe Karen *lo-bú* as 'stone-in' and *ʃɔ-bú* 'in a cave in the cliffs' as 'cliff-in.'

Eleven sampled languages, namely Hausa (one of the two relevant terms denotes "a hole in the ground where mat-makers sit and work" inter alia), Noni, Ngaanyatjarra, Badaga, Khalkha, Nivkh, Kildin Saami, Cheyenne, Arabela, Toba, and Hawaiian colexify 'cave' with 'den' or 'lair' of an animal, and in five sampled languages, Ngambay, Khalkha, Nivkh, Nez Perce (colexifying also 'bluff' and 'rockshelter') and Pipil, the relevant terms also denote a 'ravine' and/or a 'valley.'

Six sampled languages of the Americas, Haida, Oneida, Quileute, Wintu (where the relevant term *holoq* contains *holo* 'dark'), Aguaruna and Cashinahua colexify 'cave' and 'tunnel,' whereas Khalkha, Sora, Itzaj, Hawaiian, and Malagasy colexify 'cave' with 'pit.' Furthermore, three of the sampled languages, Khoekhoe, Rendille, and Kosarek Yale have terms for 'cave' (in Kosarek Yale more specifically 'cave (where bats dwell)') which can also refer to a 'ledge' or an 'overhang.' The Kosarek Yale term is *bubun*, for which compare *bubu* 'liver, carrying part, point of support.'

Other less common patterns include: Efik *aba* also means 'more, gain' and 'forty, fortieth.' Ngambay *wél* also means 'rapidly,' Dongolese Nubian *gə̄r* also 'shore, bank, waterside' and 'to crush (grain)' (in the sense 'cave' it is a loanword from Arabic), Lavukaleve colexifies 'to burst,' and Meyah *mansú* can also refer to a hollow in a tree. Muna *lia* also denotes 'stinging hair on leaves, bristles.' Rotokas *kakiua* contains *kaki* 'opened, cracked, split open' and the classifier *ua* for narrow objects. The same suffix is present in *tarieua*, where the lexical root means 'to encircle.' Sko *long* also means 'key,' and Yir Yoront colexifies 'cave' with 'jail.' Abzakh Adyghe *nəbɣe* also means 'nest,' Badaga colexifies 'place' generally, and Basque *leize* can also refer to an 'abyss,' 'chasm,' or 'depths' in general. Khalkha *kebtesi* also means 'uterus' and 'placenta.' Sora *'rupa:n* also means 'pit' and 'pool in stream' and is derived from *'rupa:-* "to form pits and hollows on account of rain." Tuscarora *yunhú·čę?* is built around the lexical root *-nhučę-* meaning 'be a corner' and also means 'cape' and 'nook,' while Wappo *hóća* also means 'sweathouse' or 'dancehouse.' San Lucas Quiaviní Zapotec *cweeb* is also the name of a children's game, and Copainalá Zoque colexifies 'hole in ground, cavity.' Aguaruna *juwaínu* can also mean 'island,' 'opening,' and 'drilled,' and Aymara *putu* also can refer to a small hole in the ground where produce is stored. Cashinahua *bean debu* contains *debu* 'point, end, source of river,' while Cavineña colexifies 'cave' (and 'hole') with 'well.' Huambisa *waa* can also refer to a 'partridge,' Ancash Quechua *machay* also means 'to get drunk,' and the Piro term for 'cave,' *mahka*, can also refer to a 'hill, ascent.' Tehuelche *ma:* is also one of the variants of a verb meaning 'to kill.' Hawaiian *ana* colexifies 'larynx' inter alia, *lua* in the same language also denotes 'grave,' 'mine,' and 'crater' alongside other meanings, and *pao*, again alongside other meanings, also '(arch of a) bridge' as well as 'to scoop out, dub out.' Kapingamarangi *lua* is also the numeral 'two.'

11. The Clearing

Representation: 32%
Motivated: 56%
Thereof Analyzable: 42.7% Thereof Colexifying: 13.3%
Thereof by Contiguity: 45.6% Thereof by Similarity: 2.7%
Recurrent associated meanings: open/clear/unobstructed, meadow/lawn/field/glade, valley, cut, gap, tree, nothing

Terms for 'clearing' frequently make reference by verbal elements to notions such as 'to open up,' 'to clear up,' or contain other lexical items meaning 'clear,' 'open,' or 'unobstructed.' Such terms are found in eleven sampled languages, namely Efik (where the relevant tem can also refer to 'openness' as a character trait), Nunggubuyu, Rotokas, Welsh, Upper Chehalis, Nez Perce, Bororo (where 'clear' is colexified with 'light, bright,' inter alia; Fijian colexifies these meaning directly with 'clearing'), Miskito, Yanomámi, as well as in Hawaiian (the term is curious: *pā-pū* also means 'fort, fortress' and is in this sense literally translated by lexicographers as 'gun enclosure.' It also means "plain, clear piece of ground; clear, unobstructed, visible, in plain sight, directly confronting," and since both constituents have very many meanings, it is hard to tell which of these are relevant for conceptualization), and Samoan. For instance, in Rotokas, the term for a 'clearing' is *koraua*, containing *korau* 'clear, unobstructed' and the classifier for narrow objects *ua*, and in Bororo *boe ali* is analyzable as 'thing clear;' Nunggubuyu *abaḻa-yarayara* 'open.area slender' also denotes the 'clear sky.' Itzaj has *jäwänka'il* /jäwän-kab'-il/ 'open-world-ABSTR' for both 'clearing' and 'plaza' (alongside *petexka'* /petex-kab'/ 'circle-world'). Moreover, Great Andamanese has *êr-tâlimare* 'place-clear.'

Four sampled languages, Efik, Kaluli, Kwoma, and Hawaiian, have terms referring to the fact that a clearing may come into being by the cutting down of trees (although 'tree' is not always present as a contiguity anchor): Efik *eritem* is derived from *tem* 'to cut down' (the term can also refer to a 'boiling, cooking,' which is due to *tem* also having the corresponding verbal semantics 'to boil, cook' inter alia), Kaluli has *i-kuwo:* 'tree-cut,' Kwoma *aka tobo me yii* 'house trunk tree cut' and *yii poy* 'cut completed,' and Hawaiian *kāhanahana*, which is related to *kahana* 'cutting.' In another two languages the terms make reference to the fact that on a clearing there is no plant cover: Baruya has *ku'maaya* /kutawɨ-maaya/ 'inside-none,' meaning 'open area, clear area, area without grass' and Carrier *hwozzai-ḳet* 'nothing-on' (this term colexifies 'glade'). In another three languages, terms utilize in some way the fact that a clearing improves visibility: Basque *argi-une* 'light-space/moment' (also indeed denoting a "moment of light" or a "moment of ludicity"), Bororo *baru-gwa* 'sky piece,' and Fijian *talāsiga*, made up of *tala* 'to clear away rubbish, transplant, load cargo' and *siga* 'day, daylight, sun;' in addition Great Andamanese term *elôtwâlnga* colexifies 'clearing in jungle' with 'view.' Seven sampled languages colexify 'clearing' with either 'meadow,' 'glade,' 'lawn,' or 'field.' These are Badaga, Carrier, Kashaya, Lake Miwok, Pawnee, Tuscarora, and Central Yup'ik. Five languages, Khalkha, Chickasaw, Itzaj, Lake Miwok and Pawnee (here by a term that can be literally translated as "flat ground place," also meaning 'flat ground, prairie,' and 'low level area,' inter alia)

have terms that colexify 'clearing' and 'valley,' 'glade,' or a similar meaning, the motivation probably being similar to that for colexification of 'forest' and 'mountain' (see also sections 26 and 39 as well as § 6.3.). Finally, Badaga and Itzaj colexify 'clearing' with 'gap.'

Other associations include: Khoekhoe *!gorab ~ !goras*, also denoting a "bare/denuded area (as when burnt down/mowed)," is derived from *!gora* 'barren, denuded, bald,' and Noni *ŋkase* consists of *kase* 'to clear a field' and the noun class prefix *ŋ-*. Buin *aarono* also can refer to the 'open sea' (in old style) or an 'open space' in general, while Baruya colexifies 'clearing' with 'bald.' Ngaanyatjarra *yurirl(pa)* is similarly used to refer to an 'open place,' but is also extended to mean 'room, space' and 'outside,' and Nunggubuyu *abaḻa* can also refer to a 'ritual clearing.' Badaga *naṭṭa ~ natta* means "common wasteland, common place used for non-agricultural purposes," and *tera* also "chance, opportunity, course." Khalkha *cabcijal* seems to contain *cabci-* meaning inter alia 'to blink' and 'to chop, to hew,' and Sora has terms that colexify the meanings 'hill,' 'clearing on a hill,' but also 'forest;' they are derived from a verb meaning 'to make a clearing on hill, cultivate on hill.' Chickasaw *kochchaafokka'* is analyzable as /kochcha' aa-fokha-'/ 'outside LOC-be.in-NMLZ,' and denotes "an open place out in the middle of nowhere" generally, including a 'valley.' Haida colexifies 'muskeg' and 'bog,' and Kiowa *'ʜt̂ɑ̨ɛ̨t'eidl* contains *'ʜ'* 'wood' and *t̂ɑ̨ɛ̨* 'be smooth;' the unknown element *t'eidl* is probably related to *t'ʜ'* 'to cut several' according to the consulted source. Itzaj *yam* may also refer to a 'groove' or 'canal,' and yet another Itzaj term, *paa*k, also means 'stack, packet.' Central Yup'ik *cangurneq* consists of the root *cangur* 'to lack symetry' and the postbase (cf. § 4.4.2) *-neq* 'area.' *Carr'ilquq ~ carr'ilqaq*, another Central Yup'ik term for 'clearing,' contains *carrir-* 'to clean' and the postbase *-quq* 'one that is.' Bororo *boe giguduru* appears to contain *boe* 'thing' and *gigudu* 'dust.' Fijian *lalama* also means 'transparent, too thin (of cloth),' Hawaiian *paia* also 'wall, side of a house,' as well as "walled in" by vegetation, and Samoan *lafo* inter alia also denotes a "plot of land (made ready for planting)."

12. The Cloud

Representation: 95%
Motivated: 41.2%
Therof Analyzable: 14.6% Thereof Colexifying: 27.4%
Thereof by Contiguity: 10.5% Thereof by Similarity: 26.6%
Recurrent associated meanings: fog/mist, sky, smoke, day, steam, rain, black/dark, cataract of eye, skin/leather, white, wind, faeces

33 sampled languages colexify 'cloud' (generally ignoring in the ensuing discussion if sources indicate that the relevant terms denote specific types of cloud unless this seems relevant) and 'fog/mist' (see § 6.3. for discussion of possible influence of environmental conditions, section 25 for analyzable terms for 'fog/mist' on the basis of 'cloud,' and Buck 1949: 65 for this patterm, which is pervasive by semantic shift in Indo-European). These are Bakueri, Ngambay, Dongolese Nubian, Yoruba, Anggor, Baruya, Buin, Kwoma (colexifying also 'white,' for which compare Bororo *baru kigadu* 'sky clear/white'), Meyah, Ngaanyatjarra, Rotokas, Waris, Kosarek Yale, Basque, Bezhta, Sora, Haida, Lesser Antillean

Creole French, Nez Perce (by a lexical affix), Xicotepec de Juárez Totonac, Arabela, Aymara, Maxakalí, Piro, Ancash and Imbabura Quechua, Tehuelche, Yanomámi, Hani, Lenakel, Manange, Sedang, and Takia (note also the similarity between Embera *hĩrará* 'cloud' and *hĩrãr'ã* 'fog, mist'). Colexification with 'smoke' is found in nine languages, Nez Perce (again by the lexical affix also colexifying 'fog'), Arabela, Bora, Cavineña (also colexifying 'smell'), Cayapa (by the term *ñivijcha*, perhaps containing *ñi* 'fire, flame, seed' and *vijcha* 'difference in height'), Maxakalí, and Tsafiki, and in Sedang by the analyzable term *kia hia* 'ghost light.weight,' colexifying also 'air;' moreover, Buin colexifies 'white cloud' and 'white smoke' (as well as 'to be smoking tobacco, be affected by smoke') specifically. Colexification with 'steam, vapor' is less frequent (attested in five languages, Anggor, Arabela, Bora, Cayapa (again by the term *ñivijcha*), and Tsafiki; note also the similarity between Koyraboro Senni *duule ~ duula* 'cloud' and *dullu* 'smoke, steam').

As for complex terms, Cashinahua has *nai kuin* 'sky smoke' for 'cloud' and Hupda has *j'ɨk kudu* 'smoke group' with the meaning "cloud, cloud of smoke/fog." A further common cross-linguistic association is that with 'sky' (see also Buck 1949: 65 for evidence from Latvian, Welsh, and Breton). It occurs by colexification in Efik (by the analyzable term *ikpa'enyöñ*, with *ik'pa* meaning 'animal skin' inter alia and *en'yöñ* 'sky, heavens;' the association with 'skin' is mirrored by colexification, also with other related meanings, in Toba, and compare also Kolyma Yukaghir *ni:nqa:r* and *qa:r* 'skin'), Lavukaleve, possibly in Nunggubuyu, Bezhta, dialectally in Sora, Upper Chehalis, Cheyenne (the meaning 'sky' is considered insecure by lexicographers in this language), Comanche, Kiowa, Lakhota (here also colexifying 'heaven'), Nez Perce, Miskito, and Rama (by the analyzable term *núnik kás* 'sun/day meat;' note in this context that Rao *grača* 'cloud' appears to contain *gra* 'sun' and that Yei *miramakər* seems to contain *makər* 'flesh'). As for complex terms, alongside the one in Cashinahua already mentioned above, Efik has *idiök' enyöñ* /idiök' en'yöñ/ 'bad/unpleasant-sky,' Yoruba *àwọ̀sánmà* /awọ̀-sánmà/ 'color/image-sky' and *ojú-sanmà* 'eye-sky' (alongside *ojú ọ̀run* 'eye heaven'), Dadibi *bulu gi dagadu* 'taro.type garden sky' (denoting an "alto-cumulus cloud formation" specifically), Ket *espul* /ēs-hul/ 'sky-mound,' Central Yup'ik *qilaggluk* /qilak-rrluk/ 'sky-one.that.has.departed.from.its.natural.state,' Bororo *baru kigadu* 'sky clear/white' (meaning 'white cloud'), Guaraní *arai* /ára-i/ 'sky/day-DIM' (*ara* also has further meanings, and the term is said to be a contraction of *ára* in the sense of 'weather' and *vai* 'bad'), and Jarawara *neme sabi* 'sky/up be.foamy.' Furthermore, the association is realized by noun class alternation in Swahili, and there is a semianalyzable term in Wayampi, which also denotes a magical plant that makes clouds come. Alongside Guaraní and Rama, which have an association between 'cloud' and 'day' by virtue of colexifying 'sky' and 'day,' there are also other cases with this pattern. Blackfoot *soksistsikó* is analyzable as /sok-iksistsikó/ 'above-be.day,' and in Hawaiian and Samoan, *ao* may refer to both 'cloud' and 'day' (in Hawaiian, also to the 'dawn' and other concepts and in Samoan also to the 'head'), but notably not to the 'sky.' Furthermore, Badaga *ka:re* is glossed as "overcast sky, clouds hovering overhead."

Occasionally, as already seen in the example from Kyaka and Bororo, terms make reference to either the white or black/dark color of clouds. This distinction is not exclu-

sively found in motivated terms; for instance, Berik has *anis* 'white cloud' and *umiyan* 'black cloud,' whereas Chayahuita *huiriro'të'* and *yararo'të'*, 'white cloud' and 'black cloud' respectively, contain the respective color terms. Further complex terms involving either constituents meaning 'black' or 'dark' include Mali *achēluinggi*, derived from *chluing* 'black,' Sahu *lobi-lobi'i*, reduplicated from *lobi'i* 'dark,' and Lesser Antillean Creole French *an mas nwe* 'in bulk dark,' while Piro colexifies 'cloud' and 'darkness' directly.

A linguistic pattern peculiar to New Guinea are terms literally translatable as 'wind faeces': Toaripi has *mea-e* 'wind/weather faeces' and Takia *tim-tae-n* 'wind-faeces-3SG.' Four sampled languages, Mbum, Nunggubuyu, Yir Yoront, and Cahuilla, colexify '(rain)cloud' and 'rain' (Yir Yoront also 'rainbow serpent;' note that this is a typical instance of actual/potential-polysemy common in Australia). Furthermore, in Khoekhoe, a term for 'cloud' in the Nama dialect, *lnanu-s*, shares the root *lnanu-* with *lnanu-b* 'rain, thunderstorm,' and the meaning is differrentiated by the alternation of the nominal suffix. Kwoma *wayi keyi* is analyzable as 'rain make.mark.' Two languages, Hausa and Khalkha, extend a term for 'cloud' metaphorically to 'cataract of the eye' (the Hausa term with this pattern denotes 'light fleecy clouds' specifically and also bears the meaning 'cuttle-fish,' while the Khalkha term may also refer to a "cloud-like pattern or design").

Further patterns include: a further Hausa term denotes 'small clouds' and "a perfectly circular calabash;" this term, *san doki*, goes back etymologically to *sawun doki* , literally 'horse's footprint.' *Girgije,* yet another Hausa term, denotes "[l]argish clouds in wet season" as well as 'to shake off,' said of garment, plumage, and other things. In Khoekhoe, *!âub ~ !âus* is derived from *!âu-* 'expect, await' (perhaps because the presence of clouds arouse the expectation of rain? In fact, *!âus* also means 'expectation, hope'). Another register-specific Khoekhoe term for 'cloud' and 'rain,' *!hūduru-e!*, is literally 'ground-mouse,' since the first drops of rain "make marks like footprints of mice." The Noni term *kembɛw* contains the noun class marker *ke-* and is otherwise identical with *mbɛw* 'beside, near,' while Dongolese Nubian *níčč(ɪ)* 'cloud, mist' is historically derived from *níǧ* 'to sew' and hence also denotes the 'action of sewing;' the extension to 'cloud, mist' is explained in the source by them apparently veiling the environment like cloth. The Kwoma term *apoduwan keyi* (*apoduwan* 'crested or crowned pigeon;' *keyi* 'make mark') denotes 'small black clouds,' Rotokas *rukuta* may contain *ruku* meaning 'hinder, hold back, block' among other things, and Sko *a* also denotes the 'blackpalm' inter alia. Sahu *samasamama* 'fleecy clouds' is apparently formed by partial reduplication from *samama* 'crocodile.' Basque *laino* figuratively may also assume the meaning 'confusion,' and Japanese *kumo* also means 'spider.' The Pawnee term *ckaʔuʔ* also means 'face,' while Tuscarora *uráʔθeh* also means 'spleen.' Another term conceptualizing clouds using an internal organ as the vehicle is found in South America, namely Cubeo, which has *oco-penibo* 'water-liver.' San Lucas Quiaviní Zapotec *x:ca'ài* seems derived from *ca'ài* 'very early in the morning' by means of the nominal possession prefix *x:-*. Aguaruna *yujagkím* is also the name of a valley, and the Bororo term *tugu* can refer to 'cloud' as well as to a 'shadow' and 'suspicion, decision' alongside other things. Huambisa *yuragkim* contains *yurag* 'fruit,' but is not further analyzable, while Maxakalí *gõy* denotes 'cloud,' 'smoke,' as well as 'flood.' The Piro term *mko* is not only used for clouds, but also for 'tufts' or 'clusters' more generally, Ancash Quechua *puyu* also means 'moth,'

while Fijian *ō* is also used figuratively with the meaning 'sign, trace' (alongside being an interjection in ceremonial settings). Hawaiian *ʻōpua* 'puffy, cumulus or billowy cloud, cloud bank' is derived from *pua* which inter alia can mean 'flower, blossom,' but also 'to issue, come forth, emerge,' said of smoke, wind, speech and colors. Rotuman *aoga* is also the gerundive form of *ao* 'to seek,' and Sedang *xok* is also used with the meaning 'garbage.'

13. The Coal

Representation: 93
Motivated: 29.3%
Thereof Analyzable: 12.8% Thereof Colexifying: 16.4%
Thereof by Contiguity: 6% Thereof by Similarity: 3.6%
Recurrent associated meanings: embers, fire, black, ashes, burn/burnt, stone, soot

Frequently, languages colexify 'coal' and 'embers,' as do Koyraboro Senni, Noni, Swahili, Yoruba, Gurindji, Kyaka, Ngaanyatjarra, Nunggubuyu, Upper Chehalis, Cheyenne, Chickasaw, Wappo, San Lucas Quiaviní Zapotec, Copainalá Zoque, Arabela, Carib, Guaraní, Toba, Wichí, Yanomámi, Kapingamarangi (by the analyzable term *ma-lala* 'RESULTATIVE-heat.up' which also denotes "wood chips made when using adze"), Manange, and White Hmong. In addition, Wayampi has the complex term *apiĩ-ngɛ* 'embers-PAST' and Kiowa has *ʼeiṗ-kʻǫ́gyH* 'live.coal-black.' In fact, it is frequently the case that terms for 'coal' make reference to its black color, as in Blackfoot *sik-óóhkotok* 'black-stone' or Biloxi *pĕ´xĕno^{n}n´ sŭpi´*, containing *pe´ti ~ pěti´ ~ pět* 'fire' and *sŭpi´ ~ sûpi ~ sŭp* 'be black.' Further languages in which one of the constituents is the color term for 'black' or means 'dark' are found in Kwoma (*keyihapa now* 'black color/earth.pigment,' which in fact also means "black paint, black pigment used to make paint" as well as "object burnt by fire" generally), Nuuchahnulth (*tumiːš* /tum-ɪ̌š/ 'dark-consume'), Cavineña (*etiru-sehueda* 'burnt-black,' which is also capable of referring to a type of tree and the black fruit of a tree), Cubeo (*ñemichĩchi* /ñemico-chĩchi/ 'black-CLASS.SCALE-LIKE.OBJECT'), and Yanomámi (*ishiishi*, reduplicated from *ishi* 'black, burnt'). For Tasmanian, Plomley (1976: 180) mentions the term *ly.hooner veene*, consisting of *ly.hooner* 'black' and *veene* 'wood;' note also the apparent lexical relationship between Cahuilla *túl̓* 'coal' and *túl-nek ~ túl-ek* 'black' and between Bwe Karen *phɛ θə̀rɔ̀* 'coal' and *khi θə́rú* 'be very dark.' In addition, Badaga and Tuscarora colexify the meanings (Badaga also with 'vegetables, curry' and "envy, jealousy, grudge"), and a semianalyzable term for 'coal' involving the word for 'black' might be present in Abzakh Adyghe.

Alongside Biloxi, terms for coals in which 'fire' acts as a contiguity anchor are found also in a number of other languages, and the respective terms are frequently metaphor-driven. Efik *u-kañ makara* is analyzable as 'NMLZ-fire european' and denotes the "charcoal used by smiths," Yoruba *ẹyin iná* is literally 'egg fire' (*ẹyin* alone can also be used to refer to 'coal'), Meyah *mah ofóm* is analyzable as 'fire ripe,' Waris *suwesumbul ~ suwumbul* as /suwe-sumbul/ 'fire-earthworm,' and Abipón *nkaate l-ahak* as 'be.burning 3SG-wood' (Macaguán colexifies 'coal' with 'firewood,' and semianalyzable terms with constituents meaning 'tree' or 'wood' are also found in Mbum and Embera). Carib *waʼto-po*, which colexifies 'embers,' is derived from *waʼto* 'fire,' Jarawara *yifo witi* is analyzable as

'fire/firewood edge/nose/button' (this term also colexifies the meanings 'brand,' 'torch,' and 'fire' itself), Rama *abung kúng* as 'fire louse,' Tsafiki *nin calá* as 'fire silver/money,' and Tetun *ahi-knaar* as 'fire-work.' There are semianalyzable terms featuring a constituent meaning 'fire' in Noni, Sko, Upper Chehalis, Bora, Guaraní, Huambisa, and Rama, and in San Lucas Quiaviní Zapotec, *gyih* is glossed as "fire: esp., coals, embers."

Alongside Blackfoot, in which the term for 'coal' contains that for 'stone,' Carrier *tsˇĕ-tîzḳen* is analyzable as 'stones-burn,' and Great Andamanese *taili châpa* consists of *taili* 'stone' and *châpa* 'firewood.' Khoekhoe *lui-ǂnom-n* is analyzable as 'stone-produce.charcoal-NMLZ,' and a semianalyzable term where one of the constituents is 'stone' is found in Huambisa. Furthermore, six sampled languages, Burarra, Kwoma, Yir Yoront, Khalkha, Jarawara, and Sedang colexify 'coal' and '(hot) ashes,' and in Yaqui *maatum* 'coal' consists of *maatu* 'ashes' and the plural marker *-(i)m* (a semianalyzable term for 'coal' containing the word for 'ashes' is also found in Bwe Karen). Two languages, Rendille and Yanomámi, colexify 'coal' and 'soot.'

Further associations include: Buli *kaali* also means 'to rub with hand,' Ngambay *kər* also means 'forest, bus,' inter alia, and *kúl* also 'to cook, prepare food.' Gurindji *kunyini* also means 'hearth,' while Muna *gheo ~ ghio* also means 'to cover with charcoal' as well as "dry and cold (of corn, coconuts etc.)." Kyaka *langa* also means 'careless, lax, impetuous,' and Toaripi *aro* is also the name of the frigate bird (which is mostly of a deep black color) as well as a kind of squid. Nez Perce *símux* is also a man's name, Wintu *ƙal* also means 'to stare, gaze, intense visual contact, glow,' 'receive intense visual impression/sensation' and 'feather,' and Central Yup'ik *qetek* is the name of the "underground tuber of the *mare's-tail plant* (Hipperus vulgaris) or *horsetail plant* (Equisetum arvense)" and also means "piece of coal, from its similar, black appearance," while Lesser Antillean Creole French *chebon ~ chabon* also means 'carbuncle.' Aguaruna *kayúshik* might contain *kayút* 'to smoke,' and the Yanomámi term *õshõhõre* also denotes the 'burnt bone of a dead person,' Bislama *kol* also means 'cold, cool,' (due to phonological collapse of Eng. *cold* and *coal*), 'sticky, adherent,' and "greeting, call (on radio)," and Hani *meiq* also 'to teach.' Hawaiian colexifies 'coal' with 'bituminous,' Lenakel *nouanamkɨlu* contains *noua* 'fruit,' and Rotuman *mahala* may also refer to 'cinders,' 'coke,' and 'clack lead.'

14. The Coast

Representation: 59%
Motivated: 53%
Thereof Analyzable: 41.2% Thereof Colexifying: 12.4%
Thereof by Contiguity: 17.2% Thereof by Similarity: 22.7%
Recurrent associated meanings: edge/end/border/point/limit, sea, water, side, river/stream, land, hem, proximity

Clearly the most frequent lexico-semantic association for the 'coast' (or 'shore,' 'beach') is that with meanings such as 'edge,' 'end,' 'border,' 'point,' or 'limit' (see Buck 1949: 31-32 for Indo-European evidence, where this association is well-attested). This may be by colexification, as in Noni, Kyaka, Bezhta (where the relevant term also means 'land' in

general, as is the case in Dongolese Nubian; there is a semianalyzable term where this is the meaning of the identifiable constituent in Rotuman), Khalkha, Sora (colexifying also 'line'), Santiago Mexquititlan Otomí, Wintu, Aguaruna, Bororo, Chayahuita, Sáliba, Toba, Tehuelche (where the relevant term *-ork'en* is derived from *-or* 'around' and also means 'body'), and Yanomámi, but occurs more frequently by morphologically complex terms, with the second constituent being 'river,' 'sea,' or more generally 'water,' as in Buin *atiga-minno* 'sea-edge' or Dadibi *áí bogabadu* /aí bogabadu/ 'water border.' Further languages with terms of this type are Yoruba (where 'edge' is colexified with 'ear'), Sahu, Toaripi (where 'point' is colexified with 'corner,' 'peak, summit, topmost point'), Basque, Khalkha (by another term than the colexifying one listed above), Laz, Chickasaw, Lesser Antillean Creole French, Pawnee, Yaqui, Bora (where 'edge' is colexified with 'side;' note in this context that Gurindji *pirntiwirti* is reduplicated from the base *pirnti* 'side,' and that Aguaruna, Bororo, and Toba colexify 'coast' and 'side'), Guaraní, Hupda, Hawaiian, Malagasy, White Hmong, and Yay. All in all, the association is found in 31 sampled languages. Variants of the association are found in Efik, where *adaña*, which also denotes a 'morning ebb' among fishermen and 'mark, average' generally is derived from *daña* 'to fix boundaries, fix extremes,' Ket, where 'land' rather than 'edge' or 'border' is the meaning of the relevant constituent (*sest baˀŋ*, analyzable as /sēs-d-baˀŋ/ 'river-POSS-land' and *ul-baŋ-t*, possibly 'water-land-NMLZ') and Wichí, where *tewuk lhip* is literally 'river part' (Blackfoot has *iisóítahtaa* /iso-niitahtaa/ 'on.a.horizontal.surface-river'). There are also structurally different complex terms in which one of the constituents is 'sea.' One such term is Khoekhoe *huri-ammi* 'sea-deep.furrow' (*ammi* also means 'beak'), another one is Huambisa *nayantsa pakari*, containing *nayants* 'sea, ocean' and *paka* 'plain.' Yir Yoront, where there is also a complex term of the redundant type *yuwl-ther* 'sea/seashore-edge,' Abzakh Adyghe, Badaga, and Yuki colexify 'coast' and 'sea,' the latter language by the complex term *ʔuk hoṭ* 'water big,' while Mali colexifies 'coast' and 'salt water.' Semianalyzable terms one of the constituents of which is 'sea' are found in Huambisa, Rama, and Piro. Furthermore, Miskito and Rama, spoken in close proximity and in contact, share terms for the 'coast' that are strikingly similar: Miskito has *kabo lama* 'sea proximity' (though *lama* also means 'breast' inter alia) and Rama *táuli shá suka* and *táuli ska ~ táuli ská*, containing *táuli* 'sea' and *súka* 'next.' Other structures are also found for complex terms on the basis of 'water': here, the cross-linguistic variation includes terms such as Kanuri *cî njî-bè* 'mouth water-of,' Carrier *yaṯhûpa*, containing *yaṯ* 'sky' and *thû* 'water' and the Kashaya terms *qʰatow*, analyzable as /ʔahqʰa=tow/ 'water=at' and *qʰami lahwal*, containing *ahqʰa* 'water' and *lahwal* 'across.' There is a semianalyzable term where the identifiable constituent is 'water' in Manange. Khalkha and Yanomámi colexify 'coast' with 'hem;' the relevant Khalkha term *købege(n) ~ købyge* appears to be derived from *købe-* 'tack, baste in sewing,' suggesting that 'hem' is the original meaning.

Other associations include: Efik *usuk* and *ikpekhe* are derived by a nominalizing prefix from the verbs *suk* 'bring down' (this term may also refer to "[c]ountries beyond, out of the river, e.g. England" and occasionally to the 'south') and *pekhe* 'divide' (this term also may refer to a 'division,' 'curtain,' and 'apartment' inter alia). The variant *gā̋r* of Dongolese Nubian *gā̋r ~ nā̋r ~ ā̋r* also means 'to crush (grain)' and 'cave' (the latter reading is

due to borrowing from Arabic). Swahili has *pwa-ni* 'ebb-LOC,' Yoruba *àgbègbè* also means 'neighborhood, vicinity' and 'region,' and Anggor colexifies 'coast' with 'sand.' Berik has *fi eyep* 'salt around' (note also Noni *ŋgem (yoo)*, derived from a verb meaning 'to surround, encircle' by a noun class prefix; *yoo* is 'water'). Burarra *gochilawa* "coastal area, on land or out on sea" is derived from *gochila* 'abdomen,' while Nunggubuyu *madhaḻag* has a narrow reading 'beach, coastline' but may more generally refer to the 'coastal region' (the same is true of Basque *kostalde*). Basque *costa* also means 'to dock, berth.' Lesser Antillean Creole French *lakòt* can also refer to "places away from the city" generally, Nez Perce *ʔallá·y* denotes "a low place, bottom of valley" generally and also "the downriver region (toward the Pacific Coast); downward; toward the Pacific Coast." Santiago Mexquititlan Otomí *ñäni* also means 'to defend.' Pawnee has the term *huukaahaaruʔ*, containing *huuka(wi)-* 'along a stream course' and *haar* 'place;' this term also means 'dry river or lake bed.' Somewhat opaque semantically is Tuscarora *kę? tiwakęhyaʔná·ʔnyeʔ*, which is based on the verbal root *-akęhyaT-* 'to extend from.' Wintu *ʔe·l* also denotes the 'edge of fence or earthlodge,' 'at the edge, marginally' and 'all over, toward,' *ƛoq* also means 'bar, rocky bar' and 'along,' while, similarly, Central Yup'ik has *canineq*, which contains *cani* 'area beside' and the postbase (see § 4.4.2.) *-neq* 'area of.' This term also has the meaning 'area.' Arabela has a semianalyzable term containing an element meaning 'lagoon, pool,' Bororo *iba* is also the name of the place outside the village where fishermen gather before hunting or fishing to evoke certain spirits and to distribute tasks. Carib colexifies 'coast' with 'wharf,' Imbabura Quechua with 'beside,' Wayampi inter alia with 'lip,' and Yanomámi with 'periphery' and 'contour.' Great Andamanese *tôtgōra* and *îgora* appear to be derived from *gōra* 'be durable,' and Hawaiian *makālae* 'beach, shore, coast near a point' contains *lae* 'point.'

15. The Dew

Representation: 77%
Motivated: 30%
Thereof Analyzable: 16.3% Thereof Colexifying: 14.1%
Thereof by Contiguity: 6.0% Thereof by Similarity: 8.5%
Recurrent associated meanings: water, fog, cold/frost, hoarfrost, urine, snow, wet/moist, star, ice, drizzle, night, wind

In eight sampled languages, Burarra, Rotokas (by the term *vusiva*, for which compare *vusi* 'burst forth, erupt, break out'?), San Mateo del Mar Huave, Quileute, Arabela, Rama, Hawaiian, and Bwe Karen, 'dew' is colexified with 'fog, mist' (in Burarra, the colexified meaning is 'heavy fog' more specifically, and Rama also colexifies 'smoke'). Similarly, Wintu colexifies 'dewy' with 'misty.' In San Mateo del Mar Huave, the relevant term colexifying 'dew' and 'fog' is *ajmiüc naquind* 'fall cold.' This is, however, not the only language in which 'dew' is lexically associated with meanings like 'cold,' 'cool' or 'frost.' Badaga, Greek, Comanche, Itzaj, Hawaiian, and Tetun colexify one of these meanings with 'dew;' in Rotuman, 'dew' is *matit toka* /matiti koka/ 'cold be.calm/be.still' (this term also means "cool night breeze or air") and Ngaanyatjarra *nyinnga kumpu* is analyzable as 'frost urine' (note also that Carrier *terhzö*, containing *terh* 'above,' also means "hoar-frost on trees," a pattern

of colexification also encountered in Buli, Cavineña, and Tehuelche). Furthermore, Tehuelche and Hawaiian colexify 'dew' with 'ice,' and Buli, Badaga, Cavineña, and again Hawaiian colexify 'dew' and 'snow.' These associations are of little surprise when one bears in mind that dew is caused by cooling down of the air which causes atmospheric water to condense. Another subsidiary explanation may be recent semantic extension of 'dew'-words to 'ice' or 'snow' in regions where these states of water do not naturally, or only seldomly occur, as is suggested in the consulted source for the colexification of 'dew' and 'snow' in Buli.

There are many sampled language where 'dew' is expressed by a complex term with one of the constituents meaning 'water.' However, there is, with the exception of Nez Perce, no language in the sample that colexifies these two meanings (the closest to this is Lenakel, which uses the same term for 'dew' and "water on the ground after rain"). Terms with such a structure include Khoekhoe *lgâ-llgam-mi ~ lgâ-lgam-mi* 'grass-water-3SG.MASC,' Abzakh Adyghe *weseps* /we-se-psə/ 'sky-??-water/liquid,' Nivkh *fi-saχ* 'bore-water,' Itzaj *p'uj-ja'* 'swarm-water,' Central Yup'ik *merr'aq* /meq-rraq/ 'fresh.water-a.little.bit.of' (this term also means 'holy water'), Miskito *diwas laya* 'wind.from.land.or.east liquid' (note also the colexification of 'wind, storm' in Kyaka and "gentle land breeze, as of West Hawai'i" in Hawaiian), Imbabura Quechua *yaku wiki* 'water drop,' Wayampi *ɨapɨ* /ɨ̃-apɨ/ 'water/river-source' (colexifying 'dew' with 'valley' and 'spring'), Yanomámi *heuheu*, presumably containing *he* 'head, upper or lower extremity' and *u* 'liquid' (an alternative term is *hemaahu*, presumably containing *maa* 'rain' alongside *he*), Bislama *wata blong naet* 'water of night' (in Muna, 'dew, expose to dew, put something out overnight to cool' is *aloma ~ alo*, with *alo* also meaning 'night, evening;' this association is also not surprising given that the air cools down in the night and thus dew is typically found in the evening, night or morning), and Tetun *mahon-been* 'shadow-liquid.' Kiliwa has a derived term, and semianalyzable terms where the identifiable constituent is 'water' are found in Kosarek Yale, Ket, and Guaraní. Moreover, there is the idiolectal term *ok-yobi-'* 'water-be.spring-NMLZ' in Chickasaw.

In five languages, namely Anggor (perhaps), Upper Chehalis, Ineseño Chumash, Kiliwa, and Hani, the term for 'dew' is lexically related to a word for 'wet' and 'moist.' In Upper Chehalis, the word for 'dew' is *sáxʷsxʷ*, reduplicated from *sáxʷ* 'wet, damp,' and in Ineseño Chumash, *spuyan* is derived from *puy* 'being moist.' In Kiliwa, the association is by colexification, and in the rest of the mentioned languages the relevant terms are semianalyzable only. There is a notable area in the Northwest Amazon region where metaphorical terms for 'dew' are found, involving either 'star' (note again that the cause of dew is condensation of atmospheric water at night), 'urine,' or both as a source concepts. In Bora, the word for 'dew' is *míícúru ní-jpa*, probably analyzable as 'star urine-CL.liquid,' in Arabela, it is *riya shaaca* 'star:PL urine,' in Hupda, it is called *wædhɔm'ăh nɔcáw* 'star saliva.' While in these particular configurations, the association is clearly an areal phenomenon on a small scale (see Epps 2007: 285 and Zamponi 2009: 590 for data from Maipure specifically, but note also that Farr 2001: 126, citing T. and C. Weber in personal communication, reports *fial ok* 'star water' for Bimin in New Guinea), the association with 'urine' is also found elsewhere: Kaluli *eleya: ba:* contains *elé* 'moon' and *ba:* 'urine' and Ngaanyatjarra *nyinnga*

kumpu, as mentioned above, is analyzable as 'frost urine.' Finally, Wintu *pʰOn* and Arabela *shushiquia* also may refer to a 'drizzle.'

Further associations include: Efik *mbara* also denotes "drops of rain hanging on the bushes after a shower" and Hausa *ra'ba* is also the name of "[a] disease of goats and sheep in which 'bags' of fluid are found in the abdomen (like highly magnified dewdrops)," inter alia. Khoekhoe *laub* may be related to *lau*, a verb meaning 'to trickle, purl' inter alia, which also yields the formally identical *laub* 'spring, fountain.' Ngambay *tàal* is also the name of the 'jackal,' the Rotokas terms *uriteva* and *vuisiva* also mean 'manna' (presumably used in Bible translations), and Sko *kú* also means 'special armband.' Yir Yoront *kal* is also a kinship term while *kalvmr* also denotes the 'galah,' also known as rose-breasted cockatoo. Badaga colexifies 'dew' also with 'hail' and 'cold season,' and Bezhta *xida* also means 'snot.' Kiowa *'ʜ'dɑ* contains *'ʜ'* 'smoke, misty rain' and *dɑ* 'to be' (and is incidentally identical segmentally with the term for 'tree, wood' and 'plant'), and Santiago Mexquititlan Otomí *xa* also means 'to mow.' San Lucas Quiaviní Zapotec *go'ohp* also means "dumb, stupid, slow-moving." Aguaruna *jegashík* appears to contain *jéga* 'house,' while Chayahuita *poroncayo'* possibly contains *cayo'* 'egg.' In Yanomámi *hemaahu*, the first constituent might be *he* 'head,' but all these associations in languages of South America seem quite dubious semantically. Hupda *paç náŋ* is analyzable as 'sky grease,' Fijian *yau* inter alia also means 'to carry, to bring,' Hawaiian colexifies 'dew' with 'to blow' (as of a breeze), 'to hit,' and other things, Bwe Karen colexifies 'gallbladder,' and Yay 'stripe,' 'design,' and 'to write' inter alia.

16. The Dust

Representation: 92%
Motivated: 54.2%
Thereof Analyzable: 12.4% Thereof Colexifying: 41.8%
Thereof by Contiguity: 5.6% Thereof by Similarity: 19.3%
Recurrent associated meanings: powder/grit, land/soil/earth, dirt/rubbish/garbage, ashes, smoke, sand, flour, steam, fog/mist, spray of sea, pollen, gray, sawdust, insect droppings

Notable are the lexical associations found between 'dust' and some aerosols. Frequently, languages colexify 'dust' and 'smoke' (found in Bezhta, Wintu, Carib, Hupda, Ancash Quechua, Bislama, where also 'cigarette' and other meanings are colexified, Fijian, Hawaiian, and Sedang, see also Buck 1949: 18 for evidence from Indo-European) or have complex terms for 'dust' on the basis of 'smoke,' a pattern found in Guaraní, Tsafiki and Tetun, where, for instance, 'dust' is *rai-bolon* 'earth-smoke.' In a subset of these languages, Carib, Guaraní, Hupda, Ancash Quechua, Tsafiki, and Fijian, an association with 'steam' is also found, due to colexification of 'smoke' and 'steam;' in addition, in Hupda, the redundant complex term *məy jʔɨk* 'house smoke/steam/fog/dust' is available to single out the 'dust'-reading of *jʔɨk*. Tsafiki also betrays an association with 'cloud' and furthermore, four sampled languages, Abzakh Adyghe, Wintu, Hupda, and Hawaiian, also have an association between 'dust' and 'fog.'

However, these are neither the only nor the most frequent cross-linguistic associations. In 26 sampled languages, associations with 'dirt,' 'rubbish,' or 'garbage' are found, by colexification in Buli, Efik, Burarra, Kwoma, Kyaka, Ngaanyatjarra, Southeastern Tasmanian, Badaga, Chickasaw, Highland Chontal, Ineseño Chumash, Kiowa, Nuuchahnulth, Oneida, Pawnee, Wintu, Central Yup'ik, Aguaruna, Arabela (colexifying 'street dirt' specifically), Aymara, Bororo, Cayapa (by a semianalyzable term containing *pe* 'faeces'), Huambisa, Hani, and Sedang. Nuuchahnulth, in addition, has the derived term *mačq-mis* 'dirty-NMLZ,' and Khalkha has the analyzable term *boγ saγ* 'filth/rubbish/sweepings egg.of.insect' with the meaning "dirt, dust, esp. in a house, refuse, rubbish," which is mirrored by a Rotokas term colexifying 'dust' and 'insect droppings.'

22 sampled languages, without the emergence of any clear areal pattern in their distribution, colexify 'ashes' and 'dust' (again, see Buck 1949: 18 for parallel Indo-European data). These are Efik, Berik, Burarra (where the relevant term is glossed as "little particles, as ashes and sand mixed where a fire has been burning" more specifically), Lavukaleve, Mali, Nunggubuyu, Rotokas, Basque, Chukchi, Biloxi, Carrier, Upper Chehalis (colexifying 'cold ash' specifically), Highland Chontal, Lesser Antillean Creole French (by the analyzable term *sann cho* 'sand fire), Lake Miwok, Oneida, Tuscarora, Wintu, Yuki, Cavineña, Hani, and Bwe Karen. Furthermore, one of the Buli terms for 'dust' is *tanbuulum* /taung-buulum/ 'sand/soil-powdered' (which also colexifies "a special kind of fine white clay (chalk?) that is eaten (sucked) by many Ghanaian people"), in Kaluli, 'dust' is *sa:-mu* 'sand-ash,' in Bislama, it is *asis blong graon* 'ashes of ground/soil,' and in Mandarin there is the redundant compound *hui1-chen2* 'ashes-dust.' Equally frequent are languages in which there is a term for 'dust' which also has the more general meaning of 'powder' or 'grit,' as in Kwoma, Kyaka, Muna, Basque, Chukchi, Greek, Kildin Saami, Welsh, Biloxi, Highland Chontal, Lesser Antillean Creole French, San Lucas Quiaviní Zapotec, Arabela, Cavineña, Cubeo, Embera, Lengua, Piro, Ancash Quechua, Toba, Malagasy, Manange, Mandarin, Rotuman, and Samoan, (and several Indo-European languages, Buck 1949: 18). Furthermore, in Maxakalí, 'dust' is *hãpkõnõn*, putatively analyzable as /hap-kõnnõn/ 'roast-powder,' Bwe Karen has *(dɛ)gəmu* '(thing-)be.powdered' and *ha phí mu* 'earth husk be.pulverized,' and White Hmong *hmoov av* 'powder soil.' The association with 'sand' is not only found in the languages mentioned so far. By colexification, it also occurs in Efik (also with 'clay'), Burarra (here the relevant term *baluk* is said to mean "'little particles, as ashes and sand mixed where a fire has been burning"), Ngaanyatjarra, Toaripi (by a semianalyzable term containing a constituent meaning 'earth'), Khalkha, Chickasaw, and Arabela (where the relevant term also means 'beach' and 'sandpit'), and Tehuelche *te:mxen* 'dust' contains *te:m* 'soil, sand, wandering dune.' Seven sampled languages, Kildin Saami, Biloxi, Tuscarora, Sáliba, Wayampi, Manange, and Mandarin, colexify 'dust' and 'flour' (Wayampi also 'spot of reddish paint on face'), while in Mbum, 'dust' is *sômn-jâl* 'flour-soil' and in Imbabura Quechua *allpa jaku* 'land flour.' As has become clear from several of the examples cited above, 'land,' 'soil,' or 'earth' are frequently involved as a contiguity anchor in complex terms for 'dust.' Further terms of this kind that do not fall in any of the categories of semantic associations discussed so far include Yir Yoront *larrpown*, which is analyzable into the words for 'ground, earth, soil dirt,' *larr*, and 'strike,' *pow*, (both con-

stituents also colexify other meanings) followed by a noun theme formative, Kiliwa *?-mat-sa?* 'DN-earth-grease/fat,' Santiago Mexquititlan Otomí *fonthai,* analyzable as /font'i-hai/ 'spread.out-earth,' and Miskito *tasba yuyu* 'earth/soil/place grain/little.particle,' while Ineseño Chumash *šupšup'* 'dust' appears to be reduplicated from *šup* 'earth.' Furthermore, Bakueri, Efik, Dongolese Nubian, Yoruba, Dadibi, Ngaanyatjarra, Southeastern Tasmanian, Nuuchahnulth, Oneida, Wintu, Arabela, Aymara, Ancash Quechua, Toba, and Hawaiian feature terms that colexify 'dust' with 'earth,' 'soil,' or specific types thereof (Yoruba also colexifies 'mud'). Arabela *joojojua* colexifies 'starch' and 'milk powder.'

Two sampled languages, Tuscarora and Yuki, associate a color term for 'gray' with 'dust,' Tuscarora by colexification and Yuki by the analyzable term *po?oṭel*, containing *po?-* 'burn' and *poṭ-* 'gray' (these terms both also colexify 'ashes'); a semianalyzable term containing an element meaning 'gray' is also found in Upper Chehalis. Languages which employ the same term for 'dust' and 'spray of sea' are Khoekhoe, Fijian, and Hawaiian (Hawaiian also colexifies 'foam' inter alia, and Fijian also "the foam at the front of a swiftly moving canoe" specifically). Three languages of Oceania, Kyaka, Kosarek Yale, and Hawaiian have terms which colexify 'dust' and 'pollen' (Kyaka *kuku* in general may refer to a "bit, flake, tiny fragment, crumb, dust, talc, powder" and also denotes a kind of tree). Finally, Kildin Saami *puenn* also denotes 'sawdust' -a parallel is found in Chayahuita- and 'metal filings' specifically.

Other associations include the following: Efik *obu* is derived from *bu* 'to rot, corrupt,' *ntöñ* also means "dimness of vision as if a mist were before the eyes" and denotes a type of spearmint, *m'bio*, derived from *bi'o* 'to cut,' also means 'chaff, trash, sweepings' and 'mote,' and *n'tan* also denotes a plant, "the flowers of which eject a fine dust when touched." Hausa *k'ura* also dialectally means "[a]ny drinking water in which nothing has been admixed" and 'urine,' Khoekhoe *tsarab* also means "dirt dust of cloud" and 'smog,' while Rendille *malálwa* can mean, alongside "fine dust found in certain places," also 'sandstorm' and 'duststorm;' 'dust cloud' is also colexified with 'dust' itself in Gurindji, and in Upper Chehalis *sƛasóqʷ* contains a root meaning 'dusty' and 'dust flying.' Dadibi *pusugu bage* may contain *bage* 'mixed group,' Kwoma *hapasen* might contain *hapa* 'bone,' while Meyah *ofóu* is peculiar in that it is glossed meaning 'egg,' 'meaning,' and 'dust,' which is likely due to homonymy. Ngaanyatjarra *kuwiri* is also used with reference to "a place where a kangaroo has lain." Rotokas has terms for 'dust' which are based on verbal roots for 'leave' and 'arrive' respectively: *kavurao*, containing *kavu* 'leave,' and *poupou* which apparently is reduplicated from *pou* 'arrive.' Kosarek Yale *lolonga* also means 'chips, chippings' and "remains of a fire or an earth-oven" and *ubdoba* also "powder on butterfly wings." Ket colexifies 'dust' with 'mould.' Welsh *lluwch* also means 'spray' and 'snowdrift.' A semianalyzable Upper Chehalis term for 'dust' colexifying 'cold ashes' specifically appears to contain a constituent meaning 'inside.' Comanche *huhkupɨ̲ ~ huukupɨ̲* appears to contain *huu ~ huupi̲ ~ huuhpi̲ ~ huh-* 'tree, wood,' while Tuscarora *awę́hę?*, alongside 'flour,' has also been extended semantically to mean 'gunpowder,' and San Lucas Quiaviní Zapotec *deheh* to 'washing detergent.' Wintu *sud,* as a verb, also means 'to cure with smoke' and 'disinfect,' and *bukul* 'ashes, dust, dirt, soil' contains *buk-* 'dark.' Central Yup'ik *caarrluk* contains *ca-* 'something' and the postbase (see § 4.4.2) *-rrluk* 'one that has departed from

its natural state.' Cashinahua *kudu* can also be used to refer to the color 'cream,' Jarawara *hobokori* appears to be derived from *hoboko*, the name of a tree species, and Rama *úng ulúng* contains *úng* 'pot' (and colexifies 'pot dirt'). Hani *gaoqbeeq* also means 'weeds.' Great Andamanese *êrl'ōtbûbut* is analyzable as /êr-l'-ôt-bûbut/ 'place-??-3SG.POSS.BODY.PART-soot,' and there is another not otherwise analyzable term containing *êr* 'place.' Hawaiian *lele-huna ~ lele-hune* "[f]ine windblown rain spray, dust, mist; to fall as fine rain" is analyzable 'fly-particle,' and *lepo*, another term for 'dirt' in the language, can also be used to refer to 'excrements' and 'silt' inter alia. Kapingamarangi *bopobo*, colexifying 'decayed' and 'mildew,' is derived from *bobo*, meaning 'rotten, decayed, old.' Tetun colexifies 'dust' with "to crumble, fall apart, smash" among other meanings, and, finally, Yay *pun*^{5} can also mean 'manure, fertilizer.'

17. The Eclipse

Representation: 41%
Motivated: 72.2%
Thereof Analyzable: 63.1% Thereof Colexifying: 9.2%
Thereof by Contiguity: 12.8% Thereof by Similarity: 56.9%
Recurrent associated meanings: sun, moon, darkness/darken, die/kill, eat, catch, be ill, disappear, numb, swallow, go into

Terms for this concept are, where data are available, in the overwhelming majority of languages expressed by morphologically complex metaphor-driven terms, with the meanings 'sun' and/or 'moon' acting most often as contiguity anchors. The most commonly recurring pattern features lexical associations with either 'to die' or 'to kill.' This is found in Chickasaw, Central Yup'ik, Bororo, Chayahuita, Sáliba, Kapingamarangi, Lenakel (where 'to die' is colexified with 'be ill;' San Lucas Quiaviní Zapotec colexifies 'be in eclipse' with 'be ill, have menstrual period,' and Ancash Quechua has *rupay qishyan* /rupay qishyay-n/ 'sun become.ill-3SG'), and Tetun by morphologically complex terms (for instance, Tetun *loro-mate* 'sun-die'), and by colexification in Central Yup'ik which has the complex term *iraluq nata-luq* 'moon/month die-??,' but also colexifies the relevant meanings in the verbal root *nala-*, which can either mean 'to die,' 'to become numb,' and 'to be eclipsed' as well as convey other meanings. Furthermore, Kiliwa has *t+hiy=h+nyuu* 'OBJ+spirit=3+kill' and Samoan *gase-toto* 'be.numb/be.dead-blood' (note the parallelism in the association between numbness and death in Central Yup'ik and Samoan).

Another pattern that is common in particular in languages of Southeast Asia is complex terms for 'eclipse' that are based on the respective words for 'eat' plus the name of an animal. Thus, in Yay, an 'eclipse' is called *baaŋ*^{2} *kun*^{1} *duan*^{1} 'flying.squirrel eat/drink month' or *baaŋ*^{2} *kun*^{1} *taaŋ*^{1}*van*^{4} 'flying.squirrel eat/drink sun;' in Sedang, a lunar eclipse is called *kau ka khêˊ* 'kau.fish eat moon' and a solar eclipse *kơxêˊ ka hài* 'centipede eat sun,' while in Hani, the 'eclipse' is called *ba'la neivq-keeq zaq* 'moon spirit-dog eat' or *naolma neivq-keeq zaq* 'sun spirit-dog eat;' the "spirit-dog" is said in the consulted source to figure in a traditional story. Furthermore, Sora and Upper Chehalis have complex terms for the eclipse involving verbs meaning 'to swallow' and 'to eat' respectively (the association with

'swallow' is diachronically detectable in Haida as well), and the Xicotepec de Juárez Totonac term *hua'can*, which is combined with the respective terms for 'sun' and 'moon' to refer to a solar or lunar eclipse respectively, contains *hua'* 'eat, drink.' Some degree of areal convergence may be assumed to be responsible for terms for the 'eclipse' found in three African languages which involve a verb meaning 'to catch' (Hausa *rana tā kama wata* 'sun 3SG do catch moon,' Mbum *séù/sèséì ŋgbánà hánà séù/sèséì* 'sun/moon catch other sun/moon' and Yoruba *ìmúṣòkùkùn*, analyzable as /ì-mú-ṣe-òkùkùn/ 'NMLZ-catch-do-darkness;' similarly Khalkha has *nara(n) barixu* and *sara barixu*, containing the respective term for 'sun' and 'moon' and *bari-* meaning 'to hold, grasp, seize, catch' inter alia). In fact, terms based on the meanings 'darkness' or 'darken,' as in Yoruba, are relatively frequent cross-linguistically, occurring either by terms of the lexical or derived type also in Khoekhoe, Dadibi, Basque, Arabela, and Fijian (for instance, Basque has *ilun-aldi* 'darkness-time' and *itzal-aldi* 'shade-time;' the first term may also refer to 'darkening' or, figuratively, 'sadness' or 'confusion,' and the second also to 'darkness' or 'a period of darkness'). Particularly noteworthy is also the Dadibi term *gilga ge begelama hulia saidao*, as it seems to be based on the idea that the sun turns over to the other side and thus does not shine anymore, the literal translation offered by lexicographers is "sun having turned over, (it) had gotten darkness." The association with 'darkness' or 'darken' is also found by colexification, namely in Cheyenne, Piro, and Hawaiian. In three sampled languages, Oneida, Cavineña and Guaraní, terms for 'eclipse' involve verbs meaning 'to disappear' and the secondary meanings 'to get lost' in Cavineña and 'flee' in Guaraní. In two languages of the Americas, Carrier and Tuscarora, the 'eclipse'-terms are based on a verb meaning 'to go into' (*sa yą-ṭe-ʻaih* 'sun sky-recess-pocket-get.in' and *θkàʻyęh* respectively which is analyzable as /či-ka-yę-h/ 'REPETITIVE-3SG.INDEF.AGENT-go.into-HAB;' note in this context also Nez Perce *ʔipnéʻculeylekse* which revolves around the verb *leylékʻ* 'to move into a hole').

Of course, there is also a number of other metaphor-driven analyzable terms for either lunar or solar eclipse in which either 'sun' or 'moon' act as contiguity anchor. These include Hausa *wata yā yi zazza'bi* 'moon 3SG.MASC do feverishness' and *wata yā yi mashasshara* 'moon 3SG.MASC do small.pox/feverishness,' Noni *diuu bo kpwee dwee lo* 'sun ?? moon bridge take.by.force' and *diuu ɛ baŋ kpwee* 'sun to cover moon' for 'lunar eclipse' and *diuu bo kpwee ɛ tasɛn e mvunsheeŋ*, which involves the words for 'sun' and 'moon' as well as *tasɛn* 'to meet' for the 'solar eclipse,' Nivkh *kʻjeṇ mu yr* 'sun boat time,' and Kaingang *mĩg tỹ kysã/rã mãn* 'panther with sun/moon carry.' Quileute *xitłítssil pitítschoʔ* is not amenable to a precise morphological analysis on the basis of the source, but the literal meaning is given as "a monster bit away chunks of the moon," and Toba *qaỹapo'oguet na l'edaxa añi nala'* contains *'edaxa* 'light, flame' and *nala'* 'sun.' In addition, Great Andamanese *lajabaginga*, to which either *bôdo* 'sun' or *ôgar* 'moon' is added to refer to 'solar eclipse' and 'lunar eclipse' respectively, might contain *jabagi* 'to damage.' Semianalyzable terms on the basis of 'sun' or 'moon' are found in Ineseño Chumash and Cavineña.

Further isolated associations include: Buli *yesinta chaab joka* contains *yesinta* 'shadows' (the literal translation offered by the lexicographer is "shadows entering each other"), while Efik *erinïme* is derived from *nïme* meaning 'to extinguish, eclipse' inter alia, and indeed can also refer to 'extinguishing,' 'extinction,' or 'vanishing.' Katcha *thigirono*

ma thɛrɛ and *thigirono ma ndhinaia* are presumably related to *thigidono* 'silence' (*thɛrɛ* is 'moon' and *ndhinaia* appears to be a variant of *ndhanaya* 'sun'). Welsh *diffyg* means 'lack' and *rhagori ar* is literally 'excel on.' Wintu *ċil* may be related to a word for 'bird' and 'bear' of the same shape. San Lucas Quiaviní Zapotec *rda'àu lohoh* contains *rda'àu* 'get shut, be stuck' and *lohoh* 'mouth, in front of, on top of.' Bora *íjyúnubááve* contains *íjyúnu* 'become night,' and there is also a semianalyzable term containing an element meaning 'be night' in Blackfoot. Lesser Antillean Creole French *éklips* is also used to refer to a 'loss of brilliance.'

18. The Egg

Representation: 97%
Motivated: 26.2%
Thereof Analyzable: 5.2% Thereof Colexifying: 21.0%
Thereof by Contiguity: 1.4% Thereof by Similarity: 1.3%
Recurrent associated meanings: testicles, seed, fruit, child, hen/chicken/bird, nut, spawn, ovum/zygote, male genitals

Terms for the 'egg' of a bird or a chicken specifically are cross-linguistically frequently associated with other smallish round objects, such as 'fruit,' 'nut,' 'seed,' or 'stone' in a mixture of analyzable and colexifying terms. In seven sampled languages, Buin, Toaripi, Kosarek Yale, Comanche, Hawaiian, Samoan, and Takia, 'egg' is lexically associated with 'fruit.' As becomes clear from the above list, this pattern is particularly frequent in languages of the broader New Guinea area (see McElhanon and Voorhoeve 1970: 29, Laycock 1970: 1141-1142). The association is mostly realized by colexification, except for Toaripi (*ori fare* 'bird fruit') and Comanche (*kokoráʔa pokopi* 'chicken fruit'). Also in New Guinea, colexification with 'nut' is found, specifically in Dadibi (indeed, the relevant term is also glossed as 'small object'), Kaluli, Kyaka, and Takia (note that Takia belongs to both groups, and indeed, it is the same lexeme that can refer to either 'egg,' 'fruit,' or 'nut'). The same areal hotspot in New Guinea is discernible for an association with 'seed,' and also here, Takia, as well as Kyaka, participate in the pattern by virtue of the same term having a wide semantic range. Other languages in which 'seed' and 'egg' are colexified are Buin, Rotokas, Kosarek Yale, Abzakh Adyghe (inter alia), Wintu (where *λu* also means to 'stab, pierce, poke'), and Hawaiian, while Baruya colexifies 'sprout,' 'shoots,' and 'seedling.' In addition, in Kiliwa, 'egg' is *xma?=yit-y* 'hen=seed-ATT.'

A pattern common in North America is to have terms for 'egg' based on 'child' and/or 'daughter/son' more specifically. This is found in Japanese (*tama-go* 'ball-child'), Chickasaw (*akankoshi',* analyzable as /akanka'-oshi'/ 'chicken-son;' this term in fact also denotes a 'chick'), Comanche (*kokoráʔa atùapʉ̲* 'chicken child'), and by colexification in Kiowa, San Lucas Quiaviní Zapotec, and, outside of the Americas, in Kosarek Yale and Hawaiian. A somewhat similar term is found in Abipón, where *tetarik l-kaoe-te* is analyzable as 'hen POSS.INDEF/3SG-work/creature-PL.' In general, the cross-linguistic evidence reveals that terms for 'egg' are sometimes morphologically complex with 'bird' acting as a contiguity anchor (a semianalyzable term is in addition found in Kemtuik). In contrast, there is

no language in the sample in which the reverse situation holds, i.e. in which the word for 'bird' is secondary to that for 'egg' (compare section 6; interestingly, unlike the situation observable for the meanings 'bee' and 'honey,' there is also no language which colexifies 'bird' and 'egg'), and this provides cross-linguistic support for the assumption that in Proto-Indo-European, the word for 'egg,' *h_aō(w)i-om* in Mallory and Adams's (2006: 143) reconstruction, is derived from that of 'bird,' *$h_a$ewei-*, and not the other way around, see Schindler (1969) for more detailed discussion.

Furthermore, Koyraboro Senni colexifies 'egg' and 'stone' (although the relevant term assumes the meaning 'egg' only in compounds), and in Santiago Mexquititlan Otomí *mädo* 'egg' contains *do* 'stone' (see Buck 1949: 256 on this association in Baltic languages). A very common and well-known metaphorical transfer pattern connects the meanings 'egg' and 'testicle' (also in evidence in Indo-European according to Buck 1949: 256). This is found in the sample in as many as 19 languages by colexification (and in nine by complex terms, see section 142), namely Efik, Ngaanyatjarra, Carrier, Highland Chontal, Itzaj, Santiago Mexquititlan Otomí, Pawnee (with slight deviations in form), Pipil (Cuisnahuat dialect), Tuscarora, Bororo, Miskito, Piro, Ancash Quechua, where the relevant term also colexifies 'scrotum' and 'rattle (snake),' Imbabura Quechua, Rama, Tsafiki, Yanomámi, Bwe Karen, and Lenakel. In One and Sedang, moreover, 'male genitals' rather than 'testicles' specifically are colexified with 'egg.'

Four languages, Efik, Kyaka, Sora, and Wintu, colexify 'egg' and 'spawn,' and two languages, Basque and Hawaiian, colexify 'egg' and 'ovum' or 'zygote.'

Other associations include: the Efik term *nsen* also denotes "a granule of any thing" as well as the 'foetus in the womb,' while *ñqua* is also "the general name of a bead." Ngambay *kàbbè* is also the name of a large kind of tree, Kaluli colexifies 'egg' with 'center,' and Kyaka *kapa* can also refer, inter alia, to the 'core of something,' a 'larva,' a 'friend,' or 'iron.' There are two further languages in the sample where the term for egg has other meanings which do not seem to be in any substantial relation with 'egg': Meyah *ofóu* also means 'meaning' and 'dust,' and Sentani *do* which also means 'man' and 'name.' The Kosarek Yale term *wana* also means 'heart' and "flower-stalk," Yir Yoront colexifies 'egg' with 'bud,' and Sora *adre:ŋ ~ arre:n* is derived from *arre:-* 'lo lay eggs' (this may also be the situation recoverable etymologically for Basque; compare also Hani *alwuv ~ hawuv* and *wuv* 'to hatch,' Yay *ʔan*^{1} *cay*^{2} CLASS.THING lay egg,' and for Jarawara *hife/hifene* 'egg' *hifa* 'brood, lay eggs;' Itzaj colexifies 'egg' and 'to lay egg.'). Abzakh Adyghe *-č*^{ʔ}*e* also means 'to grow' inter alia. Highland Chontal colexifies 'egg' with 'cocoon,' and Kiowa is unusual in apparently employing the same term for 'child,' 'egg,' and 'semen,' *ʾiH* (there is also the term *ʾih-t̂*$_{H}$*ę* 'child/egg/semen-white' to single out the 'egg'-reading), and the general meaning of the Wintu *pe·λ* in *pe·λabuhabe* is stated to be "rounded, smooth and shiny," and the term denoting 'eggs' containing it is also used to refer to 'marbles,' 'apples,' 'watermelons' and moreover means 'baldheaded.' Cashinahua *bachi* can also refer to a 'piece of clothing' as well as to a 'wasp nest' and a 'blackhead.' Cayapa *napipu* may contain *pu* 'thorn, thin bone.' Cubeo *jĩdʉ* (analyzable as *ji*, which bears the lexical semantic meaning 'egg' and the classifier *-dʉ* for roundish three-dimensional objects) can also be used to refer to any kind of protuberance in general. Maxakalí *xu'uk* has a short form *xuk*; there is a verb of that form

meaning "to carry in a bag or sack, to be pregnant." Ancash Quechua *lluntu* can also be used with reference to 'hail' (although this term is said to be specific to child language), and Toba *lco'oue'* is derived from a verb meaning 'to give birth.' The correct analysis (if any) of Tsafiki *pi'poca* is not straightforward; it might consist of *pi* 'water, liquid' and *poca* 'cane of the Guadua bamboo.' Hawaiian *hua* has a wide range of potential referents, including alongside those already discussed also 'tuber,' 'produce, yield' alongside 'round objects' in general. Kapingamarangi *ngogo* can also mean 'brain' and 'zero,' probably both because of the similarity concering the roundish appearance, while Rotuman is unique in colexifiying the meanings 'egg' and 'chrysalis.' Samoan *fua* also means 'flower, bloom' as well as 'products,' similar to the cognate Hawaiian term, and Takia *patu-* may also mean 'shell' and 'back.' Yay *cay*^{2} is also a verb meaning 'to lay eggs.'

19. The Embers

Representation: 64%
Motivated: 55.9%
Thereof Analyzable: 26.2% Thereof Colexifying: 30.2%
Thereof by Contiguity: 16.5% Thereof by Similarity: 6.9%
Recurrent associated meanings: coal, fire, flame, ashes, burn, hot/heat/warm, spark, cinders, wood/firewood, small, red, grain

Rather than being based on similarity, terms for the 'embers' (or 'live coal') are most frequently associated lexically with meanings from the same domain by contiguity, that is, meanings being related in some way to 'fire.' Six sampled languages, Kosarek Yale, Khalkha, Carrier, Kiliwa, Cubeo, and Ancash Quechua colexify 'embers' and '(hot) ashes' (Carrier also 'dust'), and in two languages, the term for 'embers' is related by word-formation to that for 'ashes' (Mandarin *hui1-jin4* 'ashes-what.is.left.from.fire' and Yoruba *eérú gbígbóná* 'ashes hot'). Still more frequent, however, is an association with 'coal,' and this pattern also occurs formally by both colexification and morphological analyzable terms. Among the colexifying languages are Koyraboro Senni, Noni, Swahili, Yoruba, Gurindji, Kyaka, Ngaanyatjarra, Nunggubuyu (the term is also glossed as "heart of fire"), Upper Chehalis, Cheyenne, Chickasaw, Wappo, San Lucas Quiaviní Zapotec, Copainalá Zoque, Arabela, Carib (by the term *wa'topo,* derived from *wa'to* 'fire'), Guaraní, Toba, Wichí, Yanomámi, Kapingamarangi (by the term *malala*, derived from the verb *lala* 'to heat up' and colexifying "wood chips made when using adze"), Manange, and White Hmong, and it may be present diachronically in Ket. Where the association is by analyzable terms, the additional constituent is often also from the same semantic field. For instance, Buli has *kaala bolim* 'coal:PL fire/light' ('fire' is also present in Lesser Antillean Creole French *ti mòso chabon difé* 'small piece coal fire,' which also has *sann cho* 'sand hot/burning'), Basque *ikatz bizi* and *ikatz gorri* 'coal live' and 'coal red,' Bororo *eradu uru ~ joradu uru* 'coal heat,' Miskito *kwasku lakni* 'coal flame,' Kaingang *pránh gru* 'coal flame/burning,' and Fijian *qilaiso sa waqa tū* 'charcoal PART be.burning stand' ('burn' is one of the additional constituents also in Sko, compare also Lesser Antillean Creole French *sann cho* 'sand hot/burning' and Abipón *leergRaie*, which consists of grammatical material that is built around the root *eerg-* 'burn,

sparkle;' this term colexifies 'nettle'). Similar to the associations with 'red,' Vietnamese has *than hồng* 'coal pink,' and in addition, there are semianalyzable terms one of the constituents of which is 'coal' in Yuki, Miskito, and Sáliba. All of the above associations also recur in different configurations. Meanings like 'heat' and 'warmth' are also relatively frequently associated in the languages of the sample, occurring alongside the cases of Yoruba, Lesser Antillean Creole French, Bororo, and Kapingamarangi that were already mentioned also in Kiliwa (*kw+pal*, which also colexifies 'ashes,' is analyzable as 'PERF+hot;' note the parallel to the Carib term where the morpheme acting to derive the complex term makes reference to the past), and Great Andamanese (*arpîjl'igûya,* presumable containing *pîj* 'hair, feather' alongside *ûya* 'warm'). A term for 'red' is alongside Basque also one of the consituent of the relevant Biloxi term, where *pĕ´xĕnoⁿni´ tcti´* contains *pe´ti ~ pĕti´ ~ pĕt* 'fire' and *tcti* 'red.' Alongside Sko, Lesser Antillean Creole French, Kaingang, and Fijian, 'burn' is also the associated meaning as revealed by analyzable terms in Japanese (by a derived term) and Bwe Karen (*mɪ-ú* 'fire-burn/catch.fire'); *tequeyashi* 'burning' is furthermore one of the constituents of Arabela *tequeyashi nootunenu* 'embers' and Wayampi *apɨ̃ĩ* contains *apɨ̃-* 'burn down.' Furthermore Hupda *teg hõ* 'wood burn' colexifies 'fire,' 'flame' and 'embers,' as does Badaga *kiccu* (the term also means 'conflagration,' 'bonfire' and, figuratively, 'jealousy' inter alia), while Buli colexifies 'embers' and 'flame' (but not 'fire'), Cashinahua and Piro colexify 'embers' and 'flame,' and Aymara also colexifies all three meanings, but by two different terms. 'Wood' or 'firewood' are also sometimes encountered as constituents in complex terms, occurring, alongside the special case of Hupda, in Efik (*ñkpri ifia ikañ* 'small:PL firewood fire') and Sora (*ərɑːdeˈtʊdən* /əˈrɑː-de-tʊd-ən/ 'wood-??-fire-N.SFX'). Four languages, Basque, Khalkha, Haida, and Tuscarora colexify the meanings 'embers' and 'spark' (in Basque, the relevant term can also mean 'spunk, pep' colloquially and in Khalkha it also means "slendour, grandeur, glory, energy" and 'spirit'). In two sampled languages of the Americas, Santiago Mexquititlan Otomí and Imbabura Quechua, there are parallel terms that involve a metaphorical transfer from 'grain' to 'embers' by analyzable terms with 'fire' acting as a contiguity anchor. The Otomí term is *d̲ẹ̲spi* /d̲ẹ̲-tsibi/ 'fire-grain' and the Imbabura Quechua one, *nina muru*, is also analyzable as 'fire grain.' As already mentioned above, 'fire' in general is unsurprisingly the most "popular" contiguity anchor cross-linguistically. Buli *bolim ngiak* might be analyzable as 'fire origin' or as 'fire liquid' (analysis is considered insecure by the lexicographer), Kosarek Yale *auk wana* is analyzable as 'fire fruit/seed/egg,' and semianalyzable terms where 'fire' is one of the constituents are found in Ngambay, Noni, Kaluli, Toaripi, Kosarek Yale, Cheyenne, Nez Perce, Pipil, Chayahuita, Guaraní, and Yay. Alongside colexification in Hupda, Badaga, and Aymara, San Lucas Quiaviní Zapotec *gyih* is glossed as "fire: esp., coals, embers." Finally, Cubeo *ũacovɯ* may contain *ũaco* 'cinder field,' and Lengua and Tetun in fact colexify 'embers' with 'cinders.'

Other associations include: Rendille is unique in using the same word for 'embers' and 'testicle,' *jiláh*. Yoruba *oguna ṣùṣu* contains *ṣùṣu* 'greatly,' and the relevant Anggor term may also refer to glowing metal. Kyaka *langa* also means 'careless, lax, impetuous.' The Mali term *sachongini* is derived from *sachon* 'eye' by means of the diminutive singular suffix *-ini*. Abzakh Adyghe *meɜ̇ˤˀeżˤekˤ* contains *żˤekˤ*, 'ember, surrounding of fire,' which in

turn contains *k°(e)* 'surrounding, depression.' Badaga *ganda* also means "sandalwood paste." Welsh *marwor* appears to be derived from *marw* 'to die,' and *marwydos*, another Welsh term, seems to be related to the verb as well. Aymara *k'aja* also denotes a 'loud color' as well as 'high temperature.' The meaning potential of Tuscarora *učí?reh* includes "candle, ember, flash of light, lamp, light, spark, taper." Central Yup'ik has *cupun ~ cup'un* /cupe-n/ 'blow-device.for' (this term also means 'straw' and 'rifle').

20. The Estuary
Representation: 32%
Motivated: 49.6%
Thereof Analyzable: 36.5% Thereof Colexifying: 15.2%
Thereof by Contiguity: 9.9% Thereof by Similarity: 27%
Recurrent associated meanings: mouth, river/stream, opening, water, foot/leg, branch, last, sea, flow out/go out

The 'estuary' is a meaning more often expressed by terms with similarity rather than contiguity as the underlying semantic relation; in either case, the relevant terms are mostly analyzable. The most frequent metaphor-driven pattern is the association with 'mouth,' occurring in eleven of the sampled languages, namely Efik, Khoekhoe, Muna, Toaripi, Abzakh Adyghe, Welsh, Upper Chehalis, Cheyenne, Nez Perce, Fijian and Mandarin; moreover, there is a semianalyzable term in Lesser Antillean Creole French. In Efik, Khoekhoe, Toaripi, Abzakh Adyghe, Upper Chehalis, Nez Perce, and Mandarin, the association is by colexification (as well as in Latin according to Buck 1949: 228), in the remaining language it is realized by morphologically complex terms. In Welsh, Muna, and Fijian, the second meaning involved is 'river,' as in Muna *wobha-no laa* 'mouth-POSS river,' and in Cheyenne, the verb *ho'ome* meaning 'be the confluence, be the mouth of a river' perhaps contains *ho'* 'arrive' and *óm* 'mouth' (in many of these languages there are also additionally colexified meanings peculiar to 'mouth,' such as 'opening' in general, 'beak,' etc., see § 6.2.3.2. for discussion). In addition, the Chayahuita term also appears to contain the word for 'mouth' and that for 'to berth,' and the Kolyma Yukaghir term *aŋil'* 'opening, river-mouth' derives diachronically from the same root as 'mouth' (note that Arabela *jiyacuaji* 'estuary, entry to path' contains *jiya* meaning 'hole, earth' inter alia). Somewhat parallel to the pattern of colexification in Kolyma Yukaghir, 'opening' is also colexified with 'estuary' and 'mouth' in Efik, Khoekhoe, Toaripi, Abzakh Adyghe, and Mandarin.

Another body-part metaphor that is however rarer is that based on 'foot' or 'leg,' which occurs in Muna and Tetun, in the former language by colexification, in the latter by the analyzable term *mota-ain* 'river-leg/foot' (note also that Haida *t'aa* also means 'foot of trail' and 'foot of bed,' compare the noun *st'a* 'foot'). Chukchi colexifies 'estuary' with 'throat,' and Wayampi with 'head' inter alia. An association not based on a body-part that is still clearly metaphorical in nature is that with 'branch' occurring in Burarra (also with 'arm, wing') and Basque. In the latter language, the term for estuary is *itsas-adar* 'sea-branch/horn,' the association with 'sea' is also present by a derived term in Welsh.

Contiguity-based analyzable terms include ones based on a verb meaning 'to flow out' or 'to go out' in Chukchi and Nuuchahnulth, and the Kashaya and Piro terms for 'estuary,' which include a constituent part with the meaning 'to fork off' or 'fork, intersection' (compare Piro *skitha* 'estuary' and *ski* 'base intersection, fork;' this term also denotes a 'confluence of rivers'). Obviously, analyzable terms frequently include constituents with the meaning 'water' or 'river' as contiguity-anchoring devices. Among these, 'river' is the most commonly used contiguous concept. Further terms of this kind include Kwoma *pa bogo* 'river junction/bend,' Sko *pá-long* 'river-hole,' Yanomámi *u paa thapi pa*, where *u* is 'liquid, river,' *thapi* is 'further' and *pa* 'away' (this term is said to mean 'above the estuary'), and Bwe Karen *lɔ̀-ká* 'river-bottom.' Moreover, Welsh colexifies 'estuary, confluence' and 'stream' directly. Terms with 'water' include Carrier *thû-ḳĕ-tcĕ* 'water-on-tail,' Kashaya *šohq^{h}awi*, analyzable as /ʔašo-ʔahq^{h}a-wi/ 'south-water-at' (note for this term that *šohq^{h}a* 'south water' is a site at the mouth of the Russian River in Northern California, so this term is more of a toponym rather than a general term for 'estuary'), as well as the apparently cognate terms of Hawaiian *muli-wai* 'after/last-water' and Samoan *muli-vai* 'be.last-water' (*muli* has other meanings in both languages, in Samoan, inter alia 'buttocks,' so here a body-part metaphor might be underlying the conceptualization).

Further associations include the following: Buin *tope* also means 'to dry out, wither,' Burarra colexifies 'mouth of river' with 'lower back, tail bone.' This term also has the more general reading of "base or rear end of anything," while the Yir Yoront term *ngopngrr* denotes the property 'wide' in a general sense, and, more particularly, a wide place in a river and an estuary. Another colexifying term is found in Khalkha, where *ada*ɣ also denotes any kind of 'end' inter alia. Lesser Antillean Creole French *labouchwi* also means 'harbor,' and Tuscarora *nyawé·kę?* is analyzable as /t-ya-ek-ę-ʔ/ 'CISLOC-3SG.INDEF.PATIENT-liquid-fall-PUNCTUAL' (presence of roots is sure, precise analysis otherwise inferred and perhaps erroneous).

21. The Feather

Representation: 95%
Motivated: 56.0%
Thereof Analyzable: 11.2% Thereof Colexifying: 45.0%
Thereof by Contiguity: 10.4% Thereof by Similarity: 40.6%
Recurrent associated meanings: hair, fur/wool, wing, leaf, bird, bristle, quill, pen, beard, needle of tree, wool, scale

Most commonly, terms for 'feather' are associated lexically with the meanings 'hair' and/or 'fur' and 'wool,' both by colexification as well as by morphologically complex terms, in which case most often 'bird' acts as a contiguity anchor. Colexification with 'hair' (without a difference made between 'body hair' and 'head hair' in the ensuing discussion) is found in as many as 41 sampled languages, namely Buli, Efik, Hausa, Dongolese Nubian, Anggor, Berik, Burarra, Kaluli, Kwoma, Kyaka, Meyah, Muna, Sahu, Sentani, Middle-Eastern and Southeastern Tasmanian, Toaripi, Kosarek Yale, Abzakh Adyghe (where the relevant term also means 'tooth' and 'seed'), Kolyma Yukaghir, Biloxi, Chickasaw, San

Mateo del Mar Huave, Central Yup'ik, Copainalá Zoque, Abipón (colexifying also 'leather'), Arabela, Chayahuita, Guaraní, Lengua, Rama, Sáliba, Toba, Fijian, Great Andamanese, Hawaiian, Kapingamarangi, Bwe Karen (which also colexifies 'between' and "to be strong, forceful"), Lenakel, Malagasy (colexifying also 'moss'), Tetun, and Yay. Colexification with 'fur' or 'wool' is found in 39 languages, namely Dongolese Nubian, Swahili, Baruya, Burarra, Kaluli, Kwoma, Kyaka, Meyah, Muna, Nunggubuyu (where the reading 'feather' is rare), Sahu, Middle-Eastern and Southeastern Tasmanian, Toaripi, Kosarek Yale, Kolyma Yukaghir, Abzakh Adyghe, Chickasaw, San Mateo del Mar Huave, Central Yup'ik, Copainalá Zoque, Aguaruna, Arabela, Bora, Chayahuita, Guaraní, Huambisa, Kaingang, Lengua, Miskito, Rama, Toba, Tsafiki, Hani, Hawaiian, Bwe Karen, Lenakel, Tetun, White Hmong, and Yay. The two categories are not mutually exclusive, since a large number of languages does not distinguish between 'hair' and 'fur/wool' lexically. There are also complex terms such as Hupda *hũtǽh pã́t* 'bird hair,' which are also found in Yei, Laz, Wichí, and Vietnamese, and thus obviously without any particular areal hotspot in a certain region of the world. Other complex terms where one of the constituents is 'hair' include Mali *chēsengvēs,* derived from *kēseng* 'hair,' Santiago Mexquititlan Otomí *xi'ni* /xi-oni/ 'body.hair-chicken' and Kapingamarangi *ngaa-hulu* 'movement-hair.' Complex terms on the basis of 'fur' or 'wool' are less frequent, and are only encountered in Chukchi (*ɣalɣa-rəɣrəɣ* 'bird-fur') and Manange (*1ŋima-2mʷi* 'bird-fur'). Due to very vague semantics of the Kosarek Yale term for 'hair'/'feather' which also includes 'beard' and 'fur' in its semantic extension, an association with 'beard' is also diagnosed in that language, as is the case in Bororo. An association with 'scale' is found in Middle-Eastern and Southeastern Tasmanian and Bororo.

Other terms with 'bird' as a contiguity anchor include Sko *táng-hó* 'bird-undress' (colexifying 'bird's tail') and Bakueri *yalí yá wŭwa* 'leaf of fowl,' as well as the similar Toaripi *ori tolo* 'bird leaf' for "long feather, from wing or tail," and there are semianalyzable terms in Mbum and Kyaka. In fact, Bakueri and Toaripi are not the only languages in which an association between 'feather' and 'leaf' is found, although they are the only ones in the sample where it is realized by an analyzable term. By colexification, the association is also found in Burarra and Ngaanyatjarra (colexifying "broad leaf" more specifically), and in seven languages of the Americas, namely Chickasaw, Ineseño Chumash, Abipón, Cashinahua (colexifying also 'silver, money'), Jarawara, Lengua, and Toba. In Kolyma Yukaghir and Chickasaw, in addition, the relevant terms also denote the needles of a coniferous tree.

In 18 sampled languages, Kanuri, Yoruba, Gurindji, Lavukaleve, Ngaanyatjarra, Badaga, Greek, Sora, Acoma, Upper Chehalis, Highland Chontal, Nez Perce, Cayapa, Jarawara, Kaingang, Macaguán, Piro, and Bwe Karen, terms which colexify 'feather' and 'wing' exist, a relation also very common in Indo-European (Buck 1949: 246). In addition, Cubeo *cave-do* 'feather' is analyzable as 'wing-CLASS.LARGE.CYLINDRICAL.SLIM.AND.ACUTE.OBJECT.' Buli, Kwoma, Bora, and Rama also colexify 'feather' with 'bristle,' and Basque, Cheyenne, Lesser Antillean Creole French, and Hawaiian also with 'quill.' Finally, three languages, Basque, Nivkh, and Lesser Antillean Creole French colexify by (at least initial) functional contiguity 'feather' and 'pen' (Nivkh by further extension also 'pointed drill'). In Buin, *paru* 'feather, plumage' is glossed also as "(used for) flower," and in Nunggubuyu *–dhabag*

'feather' can also refer to the 'flowers' of certain plants (see Laycock 1970: 1145; 1975: 228); the relevant Buin term can also refer to a "tangle-net trap for phalangers, flying-foxes, rats, and other small game."

Other associations include: Buli *kok* can also refer to a 'ghost,' Hausa *gashi* also means 'character' inter alia, and *jawaska* is also used with the meaning 'the habits and customs of a country.' Baruya features unrelated terms for feathers removed from the bird and feathers when they are on the bird. Muna *wulu* also denotes a kind of small bamboo and means 'to have a dessert' as a verb, while Ngaanyatjarra *nyarlpi* can also refer to 'playing cards.' In Yir Yoront, *marr* means 'feather' and 'wing-feather' specifically, and, by a metonymic transfer typical for the Australian language area, also denotes the 'red-winged parrot' (*Aprosmictus erythropterus*) which has a bright red (hence, "salient") wing feather. Rotokas *orupa* contains *oru* 'trim down, shave away,' Abzakh Adyghe *c(e)* may also refer to a fine substance, Basque *luma* can also refer to a 'snowflake' and *hegats* also to a 'fin' and 'eaves' (this term may be related diachronically to *hatz* 'finger'), Greek *fteró* also to a 'feather-duster' and 'mudguard,' Sora *bəˈleːdən* also to "the plume worn by the Sora-men on a turban," and Cheyenne *mée'e* also means 'collarbone.' The Kiowa term, according to the consulted source, may be related to a lexeme with the meaning 'tree, wood, stick,' and Lake Miwok *pútte* also denotes "that part of the dance headpiece which fits on top of the head." Nez Perce colexifies '(large) feather, wing' with "eagle with white-tipped wings; golden eagle, *Aquila chrysaetos*," Wintu *ƙal* also means 'to stare, gaze, intense visual contact, glow,' 'receive intense visual impression/sensation' and 'coal' and *ƙalaq* also 'feather headdress,' while Central Yup'ik *melquq* is analyzable as /meqe-quq/ 'shed.hair-one.that.is' (compare perhaps also *buuni* 'feather' in the Yahi and Northern Yana dialects of Yana with *buu* 'to shed hair,' the association is however considered unsure in the consulted source). Bororo colexifies 'feather' with 'shell' of animals and insects as well as 'casing.' Toba *laue* also means 'tower, observation point.' Yanomámi *hakorakɨ* 'wing feathers' appears to contain *hako* 'shoulder.' Finally, Bislama colexifies 'feather' with 'pubic hair,' 'grass,' 'fern,' and 'mould,' Hawaiian with 'brush' inter alia, Hani with 'to be stuck onto, to smear,' Bwe Karen with 'to spread,' and Rotuman with 'eaves.' Finally, White Hmong *plaub* also means 'four.'

22. The Flame

Representation: 74%
Motivated: 63.2%
Thereof Analyzable: 36.3% Thereof Colexifying: 27.8%
Thereof by Contiguity: 27.1% Thereof by Similarity: 26.2%
Recurrent associated meanings: fire, tongue, light, burn, embers, spark, ray/beam of light, blaze, wood, lobster, flare, candle, knife blade

Terms for the 'flame' are frequently formally complex, with 'fire' typically being the contiguity anchor. The dominant pattern cross-linguistically is metaphorical in nature and involves a transfer from 'tongue' to 'flame.' This is very frequent, and is found in all areas of the world in a total of 24 languages (Efik, Hausa, Kanuri, Katcha, Mbum, Noni, Swahili, Yoruba, Baruya, Kyaka, Toaripi, Abzakh Adyghe, Bezhta, Kildin Saami, Sora, Upper Che-

halis, Highland Chontal, Ineseño Chumash, San Mateo del Mar Huave, Fijian, Lenakel, Malagasy, White Hmong, and Takia). Canonically, the terms are of the lexical type, as in Katcha *ŋgado m-isi* 'tongue ??-fire,' with the variation that in Hausa the association is by colexification, in Lenakel, the association is ("possibly," according to the consulted source) realized by reduplication (*nam* 'tongue,' *namnam* 'flame'), and in Abzakh Adyghe, the relevant term *tx°ebz ~ tx°abze* is analyzable as /tx°(e)-bze/ 'catch-tongue' rather than 'fire-tongue.' Note also Chayahuita *pën nënërinso'*, containing alongside *pën* 'fire' *nënërin* meaning 'to flame' as well as "to stick out the tongue," as well as Lake Miwok *létaw-*, which also means "to dart the tongue (said of a snake)." Other terms with 'fire' acting as a contiguity anchor are Buli *bolim ngiak*, which might be either analyzable as 'fire origin' or 'fire liquid,' Kanuri *kánnú cámbî* /kannu cambi/ 'fire it.has.given.birth.to,' Yoruba *ọwọ́-iná* 'hand/branch-fire,' Japanese *honoo* /hi-no-ho/ 'fire-GEN-ear,' Carrier *khwen-tłek* 'fire-uses.to.dart.out,' Bora *cúújuwa péétene* /cúújuwa peéte-ne/ 'fire light-CL.thing.or.action,' Bororo *eru-gu ~ joru-gu* 'fire-liquid/blood,' Jarawara *yama hirini* 'thing catch.fire,' Miskito *pauta klauanka ba* 'fire/firewood burning DEM,' Wichí *itoj lhalh* 'fire brightness,' Hawaiian *lapa ahi* 'ridge/slope fire,' and Vietnamese *ngọn lửa* 'peak fire,' while Malagasy *lèla* in *lelàfo* 'flame' colexifies 'tongue and 'blade' (*àfo* is 'fire'). There are semianalyzable terms where the identifiable constituent is 'fire' in Kosarek Yale, Waris, and Santiago Mexquititlan Otomí. Moreover, Sedang has *pla on* for 'flame,' with *on* meaning 'fire' and *pla* inter alia "to mediate, to reconcile, to break up a fight, to come between and stop a quarrel." However, similar to the association with 'blade' above, there is also *plá* 'knife blade' which may be the constituent rather than *pla*.

In addition, Chukchi *ŋəlenɣəlet* /ŋəlet-ɣəlet/ 'burn-warm.up,' Hupda *teg hɔ̃* 'wood burn,' and Yanomámi *koã wake* /kõã wake/ 'wood red' colexify 'flame' and 'fire,' and this is also the case for simplex terms in Dongolese Nubian, Berik, Badaga, Ket, Khalkha, Comanche, Nez Perce, San Lucas Quiaviní Zapotec, Arabela, Cayapa, Tehuelche and Mandarin, so there are in total fifteen languages with colexifying terms. Alongside Chukchi and Hupda, in eight further languages, Upper Chehalis, Cheyenne, Xicotepec de Juárez Totonac, Kaingang, Miskito, Kapingamarangi, Rotuman, and Samoan, associations with a verb meaning 'to burn' are encountered, in Kaingang, Kapingamarangi, and Rotuman by colexification, in the others by morphologically complex terms: Upper Chehalis *skáwmitn ~ skáwmitn* and *skáwitn* contain *káw-* 'burn.' Cheyenne has *exo'âséotse* /éxo'asé-otse/ 'burn-become,' Xicotepec de Juárez Totonac *lamaná'* is analyzable as /lama-na'/ 'burn-AGT,' Miskito *pauta klauanka ba* as 'fire burning DEM,' and Samoan has a derived term ('burn' is also a common association in Indo-European, Buck 1949: 72). Similarly, Sahu colexifies 'high flames' specifically with 'to burn fiercely.' Thirteen sampled languages, Dongolese Nubian, Buin, Gurindji, Ngaanyatjarra, Nunggubuyu, Yir Yoront, Sora, Wintu, Abipón, Bora, Bororo, Lengua, and Miskito have terms for the meaning 'flame' which betray a lexico-semantic association with meanings such as '(fire) light' or 'to light' occurring in a mixture of colexification and analyzable terms of the derived and lexical type. Colexification occurs in Dongolese Nubian, Gurindji, Ngaanyatjarra, Nunggubuyu, Rotokas, Yir Yoront, Wintu, Abipón, Bororo, Guaraní, Lengua, and Hani (Gurindji also colexifies 'torch,' Ngaanyatjarra also 'electricity,' and Yir Yoront also 'light-source' and 'flame color;'

Bororo *eru-gu ~ joru-gu* is analyzable as 'fire-liquid/blood' as mentioned above, but colexifies 'light' in addition, and the same is true of Abipón *l-irie-Ra*, which is analyzable as 'POSS.INDEF/3SG-ignite-ABSTR'). Sora *gənʔeːm-ən ~ ˈgeːm-ən* is derived from the verb *gənʔeːm- ~ ˈgeːm-* 'to light,' and Bora has, as already mentioned, *cúújuwa péétene*, which is analyzable as /cúújuwa peéte-ne/ 'fire light-CL.thing.or.action.'

In Swahili, Tuscarora, and Bororo, 'flame' and 'ray,' 'beam of light' are colexified (and in Tuscarora in addition 'halo'). Further, Meyah *éisa* is glossed as 'flame, shine' (probably it is a verb). In four languages, colexification of 'flame' with 'spark' is found. In two of them, Abzakh Adyghe and Chayahuita, the respective terms are also formally analyzable and fall into the category of terms with constituent elements meaning 'fire' and 'tongue,' while in San Mateo del Mar Huave and Bororo the terms in question are not so analyzable. Two languages of South America, Arabela and Yanomámi, colexify 'flame' and 'candle' (and Arabela colexifies also 'match' and 'wax'). Relatedly, Wintu *sayi* also colexifies 'lantern' and 'lamp,' and Gurindji employs the same term for 'flame' and 'torch' (see Buck 1949: 72 for parallel evidence from the history of Romance). The Buli, Badaga, Aymara, Cashinahua, Hupda, and Piro terms also include 'embers' in their denotational range. Khoekhoe *ǁhabub ~ ǁkhapub* is derived from *ǁhabu* 'to catch alight,' and 'flame, flare up.' The latter meaning is also colexified in Muna, and there is a semianalyzable term in Rotokas. Lesser Antillean Creole French, Pawnee, and Hani colexify 'blaze' (Hani also 'ardor'). In Hawaiian and Kapingamarangi, relevant terms also can refer to species of lobsters.

Other associations include: Dongolese Nubain colexifies 'flame' with 'flower, blossom.' The Kyaka term for 'flame' varies in form between *uu lenge, wii lenge and wilenge*. All of these terms apparently contain the Kyaka word for 'eye,' *lenge* (which, however, also means 'node, knuckle' inter alia). Sahu *lejanga* can also refer to a 'glow' inter alia. Badaga *kiccu* is also used figuratively with the meaning 'jealousy,' and *koḷḷi* is glossed as "'thicket, bush, firewood; flame, brand, firebrand," Basque *sugar* may also refer to a 'male snake,' and *lama* also means 'glare, gleam' alongside -accidentally- 'lama,' a spiritual teacher in Buddhism. Greek *flóga* also has the figurative meaning 'fire, passion,' Khalkha *zali* can also mean 'spirit' and refer to "[r]use, craft, cunning, trick, deceit," and *døly* can also mean 'to be timid, not dare.' Itzaj *jom* also means 'to break hole, perforate,' Nez Perce *ʔáˑla* also 'hell,' and Wintu *ċul* may be related to *ċul* 'pour, spill.' The Chayahuita term *ohuica pochin nininsoˈ* is apparently based on a metaphorical connection with sheep (*ohoica*), which flames are apparently felt to be (*ninin*) like (*pochin*), and for Embera *ne eráadrɨ̀*, compare *eraadrɨ́* 'lightning.' 'Flame' and 'lightning' are indeed colexified in Tsafiki. Guaraní *tendy* also means 'saliva.' Toba *l'edaxa* colexifies 'brilliance, gleaming' alongside 'heat, fever,' and similarly, Tehuelche colexifies flame with 'flickering.' Tsafiki *pinda* also means 'thunderbolt.' Hani *miqlaol* contains *laol*, a verb meaning 'be warm or hot,' and *miqbia* contains *bia* 'bright, shining, flash.' Great Andamanese *archâl* is derived from *châl* 'beam, shine.'

23. The Flood

Representation: 49%
Motivated: 49.6%
Thereof Analyzable: 30.6% Thereof Colexifying: 18.5%
Thereof by Contiguity: 28.1% Thereof by Similarity: 10.2%
Recurrent associated meanings: water, river/stream, torrent, overflow, come/move/go inland, big, current, flow, push underwater, cover, sea, swell, (be) full, scrape

Terms for 'flood' (or 'floodwater,' 'deluge,' 'flood tide') betray a variety of lexico-semantic associations cross-linguistically. Frequently, the meanings 'water' or 'sea' act as contiguity anchors in morphologically complex terms, and in four further languages, Berik, Yir Yoront (where an etymological connection with 'sand' may be present), Comanche, and Piro, semianalyzable terms are found where one of the constituents with the meaning 'water' is identifiable.

In five sampled languages, Kyaka, One, Upper Chehalis (in One and Upper Chehalis there is a single term for 'water' and 'river,' and hence, there is also a lexical association with 'river;' similarly, Ngaanyatjarra and Badaga colexify 'floodwater' with 'river' or 'stream,' and Ineseño Chumash 'stream, creek' with 'flood' itself), Chickasaw and Lake Miwok, the second element in their complex terms for 'flood' is 'big' or 'to be big.' For instance, Lake Miwok has *ʔudíikik* /ʔudí-kik/ 'be.huge-water.' Note also Efik *a'qua i'nyañ* 'great wide.expanse.of.water' and the formally redundant Central Yup'ik term *ule-rpak* 'flood/high.tide-big.' In five languages, Dadibi, Sko, Nivkh, Pawnee, and Hawaiian, the second term has the meaning of 'come,' 'go inland' or 'move forward' (for instance Dadibi *asobo wé* 'come water,' and Sko *tí hoe toe* /tí hóe toe/ 'sea go.beachwards come.up'). In the Pawnee term *kicuutaʔa*, an additional element with the meaning 'swell' is encountered; the term is analyzable as /kic-huutaʔat/ 'be.liquid-swell' (compare colexification of 'to flood' and 'to swell' in Jarawara). Similarly, Haida has *gay-hll* 'ACTION.OF.WAVES-move' (there is another term featuring *gay*). In Yoruba, there are the terms *ì-kún omi* 'NMLZ-be.full water' and *ì-ṣàn omi* 'NMLZ-flow water.' The former association with 'full' is mirrored in Rotuman by colexification (where the relevant term also has many other meanings), and the latter is also encountered by derivation in Samoan, in Khoekhoe, where *dâus*, a term for the Biblical flood, is derived from the root *dâu*, meaning 'flow, stream, run,' and furthermore, in Nuuchahnulth, where *cuupšiƛ* is analyzable as /cŭ̓p-šiƛ/ 'for.liquid.to.flow.out-MOMENTANEOUS.' In this context, note also the similarity between Rendille *'dúley* 'flood' and *'dula*, meaning inter alia 'flow, flood, be flooded,' as well as that between Yuki *ṭal-* 'flood' and *ta·-* 'to flow.' Other analyzable terms with 'water' or 'sea' as contiguity anchors include Koyraboro Senni *harihurey* 'high flood (in annual cycle),' which contains *hari* 'water, liquid' and *huru* 'to enter,' Baruya *aalya-aka* analyzable as 'water-white' and colexifying 'muddy water,' Ket *áʁul* (analyzable as /aq-ūl/ 'rot-water' and also meaning 'hole in ice'), and Cheyenne *mâhóovátó*, which is said to mean 'all over water' literally. Kashaya has *ʔahqʰa co·biʔ* (/ʔahqʰa ʰco-Xiˆᵒbiᵒc/ 'water carry-up'), Guaraní *y-sẽ* 'water-leave,' and Hawaiian *wai-holomoku* 'water-rush.' Somewhat unclear and potentially spurious associations with 'sea' acting as contiguity anchor include San Mateo del Mar Huave *tenguial ndec*, where

ndec is 'sea' and *tenguial* is a demonstrative element and Rama *táuli alka bángi* 'salt/sea sun/hot let.us' (One also has a term colexifying 'floodwaters' with 'sea,' and there is a semianalyzable term where the identifiable constituent means 'sea' in Miskito). In addition, as mentioned above, three languages in the sample, Ngaanyatjarra, Badaga, and Ineseño Chumash, colexify 'flood' or 'floodwater' and 'river, stream' (Badaga also 'ditch,' 'lake,' and 'bottomland' inter alia), while in Tetun, one term for 'flood' is *mota-tun* 'river-down/descend,' and there is a semianalyzable term where the identifiable constituent means 'water, river, lake' in Berik and one where it means 'river' only in Rama. In two languages of New Guinea, Kwoma and Kosarek Yale, there are verbs colexifying the meanings 'to cover' and 'to flood' (Kosarek Yale also colexifying 'to hide' and 'to keep secret'), and in Tuscarora, the term for 'flood' consists of the contiguity anchor 'water' and a stative verb 'cover' with the term for 'land' incorporated (*à·wę? waʔwnawérhę*, analyzable as /à·wę? w-aʔwT-awerhu-ę/ 'water 3SG.NEUT.AGENT-land-cover-STAT'). Also in Kwoma and Kosarek Yale, there are verbs meaning 'to push underwater,' 'to put under surface of water' and 'to flood;' Bislama has *draonem* (< Engl. *drown*) with the same semantic structure, and, given the language's relative proximity, it is intriguing to speculate whether this is a more general semantic pattern of the broader New Guinea area which Bislama has acquired by relexification. Similarly, 'flood' and '(to) torrent' are colexified in Kwoma, Kyaka, Basque, Guaraní, Hawaiian, and Tetun, and 'flood' and '(to) overflow' are in Buli, Muna, Comanche, Nuuchahnulth, and Hawaiian. In two languages of Oceania, Mali and Fijian, complex terms for 'flood' revolve around a verbal element meaning 'to scrape' (see Stebbins n.d. for details of the underlying conceptualization in the former). In Buin, Kwoma, and Welsh, associations with the meaning 'current' are encountered (the relevant Welsh term *llif* also means 'saw'). In Kwoma, *ukwi kiya* is analyzable as 'water.current carry,' and in the other languages the association is by colexification (Buin also colexifies "be light brown (the colour of a river in flood)").

Other associations include: Buli *mobi*, used verbally, means 'to cut,' 'to crack,' and 'to burst' (said of banks) inter alia, Hausa colexifies "ingress of mass of water" with "watersprout, cloudburst" as well as 'congestion.' Kwoma has two colexifying terms, namely *dabu*, which also means 'to pour down' or 'to spill' inter alia, and the aforementioned *ukwi kiya*, which means 'for a river to rise' and again 'to pour down.' Meyah colexifies 'flood' with 'surplus' and Kosarek Yale *yamak-* also means "to fill the air with a smell." Basque *uholde* can also be used metaphorically to a plethora of something (as can English *flood*). The relevant Chukchi term is semianalyzable, containing an element meaning 'deep,' and Yana *ʒuu-* also means 'to push' and 'to poke, to spear.' Bislama *draonem* also means 'to post a letter.' Hani *eelpuv puv* contains *puv* meaning inter alia 'to roll, to cross over,' *eeltaoq taoq* contains *taoq* 'to pound, to butt, to rub against,' and *eelpuv leiq* also contains *leiq* meaning 'to look for, to deviate' inter alia. Kapingamarangi *doloo* also denotes a species of duck. Samoan *lolo* is a nominalization of a verb meaning 'for water to run,' and Tetun *nabeen* can either mean 'to liquify' or 'to flood.'

24. The Foam

Representation: 85%
Motivated: 32.5%
Thereof Analyzable: 11.5% Thereof Colexifying: 21%
Thereof by Contiguity: 13.6% Thereof by Similarity: 13.9%
Recurrent associated meanings: bubbles, saliva/spittle, suds, water, lungs, soap, boil, spray, bud, fog, foam on mouth, swell

The meaning 'foam' ('froth,' 'scum'), when terms for it are lexically motivated, is expressed to about equal proportion by terms with an underlying semantic relation of contiguity and of similarity, and this is reflected in the two cross-linguistically most frequently associated meanings: 'bubbles' by contiguity and 'saliva, spittle' by similarity. The association with 'bubbles' occurs in fourteen languages by colexification, Mbum, Gurindji (colexifying also 'steam'), Kwoma, Nunggubuyu, Ineseño Chumash, Haida, Itzaj, Cayapa, Embera, Guaraní, Wayampi, Hawaiian, Mandarin, and Rotuman, and one language, White Hmong, has the analyzable term *npuas dej* 'bubble water,' which, however, denotes 'bubbles' itself as well. Noteworthy is the Guaraní term *tyjúi*, perhaps analyzable as /ty-jýi/ 'urine-rainbow,' which then would make reference to the iridescent reflections of the light occurring especially with bubbles on soapy water (in fact, an association with 'soap' or 'soap powder' is found in Ngaanyatjarra, Sko, and Miskito, which latter has the analyzable term *sōp laya* 'soap water,' and 'suds' is colexified in Koyraboro Senni, Burarra, Khalkha, Cahuilla, Pawnee, Aymara, Hawaiian, and Rotuman, while Nez Perce *tíˀpip* is related to *tipíˀpi* 'become sudsy').

Interesting is the association with 'saliva, spittle' or sometimes 'slave, drivel' because it is particularly frequent in New Guinea and Oceania. It is found in a total of thirteen languages, Baruya, Buin, Kwoma, Lavukaleve, Muna, Ngaanyatjarra, Sko, Southeastern and Western Tasmanian, Badaga, Lesser Antillean Creole French, Tsafiki, Bislama, Lenakel, by colexification in nine languages and by analyzable terms in four (Baruya has *aali-maagwala* 'water-saliva,' Muna *bura-no tehi* 'foam/froth-POSS sea,' where the example in the source suggests that *bura* is used for 'foam on mouth,' but this case remains somewhat unclear, Sko *fóefóe*, reduplicated from *fóe* 'spittle,' and Lenakel *nihi-noua tehe* 'liquid-mouth sea,' where *nini-noua-* 'liquid-mouth-' is 'saliva'). Similarly, Basque and Malagasy colexify 'foam on mouth' and Kaluli "froth from mouth during seizure" specifically. Given that 'saliva' has a foamy structure as well, and is also in a contiguous relationship with 'foam at the mouth' one may wonder whether in the colexifying languages, the target meaning is really 'foam' as found on water. Indeed, morphologically complex terms like that found in Lenakel suggest that this may be so (although, to be sure, each language is different and there may be some spurious cases).

In Buin, Kaluli, and Toaripi, 'foam' is colexified with 'lungs' (due to the spongy appearance of this organ), and Kaingang *jẽngéj* is glossed as 'foam of lung' (it is also common cross-linguistically to have complex terms for 'lungs' on the basis of 'foam,' see section 122). Two languages, Khoekhoe and Ancash Quechua, colexify the meanings 'foam' and 'bud' (in Khoekhoe, especially the bud of *Acacia Watkins*), and another two, Cahuilla

and Hawaiian, colexify 'foam' and 'fog.' Similarly, Ngambay colexifies 'fog' and Gurindji 'steam.' Also in two languages, an association with a verb meaning 'to boil' is found. In Itzaj, the same term may be used as a verb with the meaning 'to boil' and as a noun with the meaning 'foam,' and in Abipón, *l-apa-Ra* 'steam of boiling water, foam' is analyzable as 'POSS.INDEF/3SG-boil-ABSTR.' Efik *ëfut'* (probably accidentally also meaning 'fifteen, fifteenth') is derived from a verb colexifying 'to boil' with 'to swell, ferment, foam' and other meanings, and similarly, Khalkha *køgesy(n)* is a resultative nominalization of the verb *køge-* 'to swell, distend, intumesce, foam.' In addition, the Xicotepec de Juárez Totonac term *puput* 'foam' appears to be derived from *pupú* 'to boil,' and lexical connections between 'foam' and 'to boil' are also reported for Wintu and, diachronically, for Haida. Furthermore, Burarra and Wayampi colexify 'foam' with 'spray' on waves, and in Great Andamanese, the same root *bôag* yields the meaning 'foam on mouth' and 'foam on sea' depending on the possessive prefix attached.

Other associations include: Buli *puuk* also means 'stomach, belly' and 'pregnancy' inter alia, Hausa *kumfa* also 'and then, thereupon,' and Burarra *munjanachana* is derived from *janachana* 'make crumbs' by prefixation of the class marker *mun-*. Rotokas *ruiruiso* appears to contain *ruirui*, meaning 'to dry up' or 'to wipe up,' Nunggubuyu colexifies "flood debris," and Toaripi *ma sese* 'slight foam on waves, white caps' is analyzable as 'water fibres' (the Northern dialect of Ngaanyatjarra colexifies 'foam' with 'water' directly, there is a semianalyzable term with an element meaning 'water, river' in One, and a Ket term may be diachronically relatable to one meaning 'water' as well). Abzakh Adyghe *txʷəmbe* (with the variant *txʷərbe*) is analyzable as /txʷə-m-be/ 'white-RELAT/EPEN-hollow.' Greek is the only language in the sample where the term for 'foam,' *afrós*, can also mean 'cream,' Japanese *awa* also means 'foxtail millet,' Khalkha *ceger* also 'taboo, prohibition, abstinence' and 'quarantine,' and Sora *bub'bu·da:n* also "[b]eads of various kinds made of glass or metal, used as ornaments." Kolyma Yukaghir colexifies "dirty water which remains after cleaning a fish." Itzaj *otz'tik* and *ootz'* can also mean 'wrinkle,' Lake Miwok *póṭa* also 'to be gray, to be cloudy' and 'semen,' Tuscarora colexifies 'effervescence' with 'thin foam,' Cavineña colexifies 'tree,' and Cayapa *chimbijpu* (or *shimbijpu*) also denotes the 'bladder' of an animal. Ancash Quechua *ñawi* also means 'eye,' 'spring of water,' and has other related readings. Rama *ngú aya* and *ingu aya* literally mean 'house corn.' The Toba term for 'foam' varies in form between *lchi* and *lcochi*; *lchi* also means 'leg, tributary.' Fijian *vuso* generally also means 'top end' and 'to squeeze out the juice from leaves, coffee powder.' Hawaiian *huʻa* is, presumably by metaphor, extended to 'border' and 'suburb' among other meanings, and *ʻehu* also means 'dust,' 'pollen,' and 'faint, difficult to see' inter alia. The complex Manange term *2kju 1atsaŋpʌ 1mo 1mu* contains *2kju* 'water,' *tsaŋ* 'clean' and a negative marker inter alia, and Mandarin colexifies 'foam, bubble' with 'blister' as well as 'to soak.' Tetun has *furi-n* 'sprinkle-SINGULATIVE,' and Samoan *piapia* is reduplicated from *pia* 'secretion of the genital organs, smegma.' Bislama *nus* (< Engl. *nose*) means 'nose,' 'mucus of nose' (by metonymy) and 'foam' (by metaphor).

25. The Fog

Representation: 89%
Motivated: 47.7%
Thereof Analyzable: 13.4% Thereof Colexifying: 34.5%
Thereof by Contiguity: 1.6% Thereof by Similarity: 35.7%
Recurrent associated meanings: cloud, smoke, steam, dew, dust, cover, white, cold/frost, darkness/darkening, foam, rain, fall, land/ground, water

For the meaning 'fog,' associations with other aerosols, namely 'smoke,' 'steam,' and 'cloud' abound. These associations are frequently realized formally by colexification, but also quite often by morphologically complex terms. The interesting fact, however, is that the relationship between the meanings as revealed by analyzable items is unidirectional, in that there are complex terms for 'fog' on the basis of the aforementioned meanings, but the reverse situation is unattested in the sample.

The association with 'cloud' is present by colexification in Bakueri, Buin, Ngambay, Dongolese Nubian, Yoruba, Anggor, Baruya, Kwoma, Meyah, Ngaanyatjarra, Rotokas, Waris, Kosarek Yale, Basque, Bezhta, Sora, Haida, Lesser Antillean Creole French, Nez Perce (by a lexical affix), Xicotepec de Juárez Totonac, Arabela, Aymara, Maxakalí, Piro, Ancash and Imbabura Quechua, Tehuelche, Yanomámi, Manange, Sedang, Takia (and perhaps Tasmanian, Plomley 1976: 230, as well as perhaps Embera, where the relevant terms are almost identical segmentally). In Ngaanyatjarra, Sedang, and Takia, the colexified meaning is 'low cloud' more specifically, while even more specialized types of cloud are colexified in Buin ('white cloud'), Aymara ("fairly dark cloud"), Chayahuita ('white cloud;' there is also an analyzable term of the redundant type with the additional constituent meaning 'to rise up'), Hani ("thick cloud"), and Hawaiian ("light cloud on mountain"). Similarly, Badaga colexifies 'veil of cloud.'

Analyzable terms include Kyaka *yangama kopa* and *yuu kupa* 'morning cloud' and 'ground cloud' (both denoting 'ground fog' specifically), Upper Chehalis *s-ṗə́ɬ=šq* 'CONTINUATIVE-drop/fall=cloud,' Highland Chontal *dedíhima lummaway* 'encircle cloud,' Santiago Mexquititlan Otomí *'bo̲ngui* /'bo̲ni-gui/ 'be.stretched.out-cloud,' White Hmong *pos huab* 'moist cloud,' and Samoan *pu-ao* 'hole-cloud' (see Buck 1949: 66 for the association between 'fog' and 'cloud' in Indo-European, which is common there). In Khoekhoe, there is a semianalyzable term where the identifiable constituent is 'stratus cloud.'

'Fog' and 'smoke' are colexified in Efik (by the analyzable term *nsuñ'ikañ* /n-suñ-ikañ'/ 'soft/gentle-fire'), Buin ("white smoke" specifically is colexified here), Burarra, Gurindji (colexifying 'light fog' with 'smoke haze'), Yir Yoront, Abzakh Adyghe, Nez Perce (by a lexical affix), Wintu, Central Yup'ik (colexifying also 'dust in air'), Hupda, Jarawara, Lengua, Maxakalí, Miskito, Wayampi (by the analyzable term *atãsĩ* /ata-sĩ/ 'fire-whiteness'), and Mandarin (colexifying also 'cigarette, tobacco'). There are semianalyzable terms in Ngaanyatjarra (here, the term also denotes a "grey-leaved acacia shrub"), One, San Lucas Quiaviní Zapotec, Lengua, and Rama. Similarly, Burarra has another term colexifying 'light fog' with 'smoke screen.' Analyzable terms are Santiago Mexquititlan Otomí

'bipa, analyzable as /'bifi-pa/ 'smoke-heat,' and Arabela *cohuaja najaca* as 'white cloud/smoke/vapor.'

Finally, 'steam' (or 'vapor' generally) is colexified in Efik (again by the analyzable term *nsuñ'ikañ* /n-suñ-ikañ'/ 'soft/gentle-fire'), Anggor, Burarra, Kyaka (also colexifying 'pawpaw'), Yir Yoront, Abzakh Adyghe, Greek, Khalkha, Welsh, Wintu (colexifying also 'gas' and 'lungs'), Arabela, Hupda, Lengua, Miskito, Hawaiian, and Rotuman, which also colexifies 'fine spray' (see also Buck 1949: 67 for evidence from Swedish and Norwegian). Perhaps similarly, Sora colexifies 'black vapour' more specifically. There are two languages with analyzable terms: Ket has *ulij* /ūl-ii/ 'water-vapor' and Lesser Antillean Creole French *vape ki ka kouve late ében lanme*, containing *vape* 'steam,' *kouve* 'cover,' and *lanme* 'sea.' As discussed in some more detail in § 6.2.2.2., in the colexifying languages, sometimes more than one of the associated meanings is expressed by the same term.

Furthermore, a lexical association between 'fog' and 'dew' is found in eleven languages of the sample (see also Buck 1949: 66 for evidence from Greek): Burarra, Rotokas, Biloxi, San Mateo del Mar Huave, Quileute, Wintu, Arabela, Rama, Hawaiian (where the relevant term is also the name of a "gentle land breeze, as of West Hawai'i"), Bwe Karen and Vietnamese. The association is mostly by colexification (in Burarra, the meaning colexified is more precisely 'heavy fog,' and in Wintu, 'dewy' is colexified with 'foggy'), with the exception of Vietnamese, where 'fog' is *sương mù* 'dew blind' and Biloxi, which has *ayuxka´*, presumably /ayu-ka/ 'dew-something.' The reasons for this association remain somewhat unclear, although a likely source for it is that both 'fog' and 'dew' require low temperatures to occur, which would then be a case of motivation by temporal contiguity. Some circumstantial support comes from the San Mateo del Mar Huave term *ajmiüc naquind*, which, in addition to colexifying 'fog,' is also analyzable as 'fall cold' (compare the Upper Chehalis term involving a constituent meaning 'to drop, fall'). In fact, in two further sampled languages, Yuki and Bislama, the meaning 'fog' is expressed by morphologically complex terms associating it with 'cold' or 'frost' (which is the cause for the phenomenon in the first place). Yuki *ƙo· nạ·ṭ* contains or is related to *ƙoh* 'frost' and *nạ·ṭ* 'ice,' while Bislama *kolkol* (the reduplication base being *kol* 'cold') can itself also mean 'cold,' 'cool,' 'stale' (said of food) and other things as well, and 'be foggy' is moreover colexified with 'be frosty' in Blackfoot. An association between the color 'white' and the meaning 'fog' is found in four sampled languages, namely Kwoma (here by colexification) as well as Chickasaw, Arabela, and Wayampi. Chickasaw *tooboklhili ~ tohboklhili* is analyzable as /tohbi-oklhili/ 'be.white/be.pale-night,' Arabela *cohuaja najaca* as 'white cloud/smoke/vapour,' and Wayampi *atãsĩ*, as already mentioned above, as /ata-sĩ/ 'fire-whiteness' (the Wayampi term, as suggested by the semantics of its constituents, also colexifies 'smoke,' so it is dubious whether the association of 'white' with 'fog' should be considered as genuine in this case). Note also Rendille *'duubát* 'fog' and *dúub* "white circular hat, turban of white cloth."

Complex terms for the meaning which involve a verbal element with the meaning 'to cover' are found in four sampled languages, Cheyenne, Kiliwa, Lesser Antillean Creole French, and Hawaiian. However, the conceptualization seems to differ to some extent in spite of this commonality: in Cheyenne, *nêhpóemáno'e* can be literally translated according

to lexicographers as "closed.in(covered)-environment," whereas in Kiliwa, *yuw=hi?* is analyzable as 'eye=cover.' Hawaiian has *uhi-wai* 'covering-water' (colexifying also a "kind of tapa"); for the Lesser Antillean Creole French evidence see above. In Kolyma Yukaghir, in addition, *šažil'* ~ *šaril'* can refer both to 'fog' and a 'covering,' and indeed is derived from the root *šar-* "to cover, to bury, to press; to overtake," and can, presumably due to the sense 'bury,' also refer to a 'root.' Two languages of Eurasia, Khalkha and Welsh, colexify 'fog' and 'darkness,' 'dimness' or 'darkening,' featuring at least one term (alongside others) that can express both meanings. Similarly, Nez Perce has a word for 'fog' as well as "darkening from fog or approaching storm" which contains the lexical affix *ʔipé-* 'pertaining to smoke, clouds, fog' and is hence formally redundant, but contains a further constituent meaning 'to be dark.' Wayampi *tɛwilakɨ* contains *tɛwi* 'buttocks, rump' and *kɨ* 'to rain,' Dadibi colexifies 'mist' and 'slight rain,' and a semianalyzable term one of the consituents of which is 'rain' is also found in Noni. Relatedly, associations with 'water' acting as a contiguity anchor are found in Ket and Hawaiian, as mentioned already, and there is a semianalyzable term in Itzaj. In two languages of the sample, Cahuilla and Hawaiian, 'fog, mist' and 'foam' are colexified (Cahuilla also has another term related to a verb meaning 'to enter'). In four languages, Abzakh Adyghe, Wintu, Hupda, and Hawaiian 'dust(y)' and 'fog(gy)' are colexified (although it should be noted that the semantic range of the Hupda term is quite large, also colexifying 'smoke' and 'steam'), and Itzaj has *aj-tanat-ja'* 'MASC-dust-water.' Similarly, Central Yup'ik colexifies 'dust in air' specifically with 'haze.'

Other patterns include: Dongolese Nubian *níčč(ɪ)* is related to *níǧ* 'to sew' and also denotes the 'action of sewing;' the extension to 'cloud, mist' is explainable according to the source by clouds and fog "veiling" the environment like what one sews, that is, cloth. The Rotokas terms *guiguisiva* and *vusiva* may be related to the verbs *guiguisi* 'spray out' and *vusi* 'burst forth' respectively, compare the apparent lexical relationship between Dadibi *segeni ebo* 'mist, slight rain' and *sege ebo* 'heavy.' Muna *gawu*, as a verb, also means 'be hazy, dim, distant' said of objects or vision, Western Tasmanian perhaps colexifies 'fog' with 'shadow,' while Yir Yoront also colexifies 'spray on waves,' and similarly, Hawaiian 'rain spray.' Abzakh Adyghe *-ɣ°e-* also means 'yellow, blond' inter alia, and Basque *laino* is also used metaphorically with the meaning 'confusion.' Japanese *kiri* also means 'awl' (though *kiri* 'fog' is a nominalization of *kir* 'become foggy,' so the similarity is accidental). A precise morphological analysis of Cheyenne *nêhpoése ma'ëno* is not possible; however the literal meaning of this term according to the lexicographers is 'the turtle is shrouded.' Guaraní *tatatîna* contains *tata* 'fire.' Ancash Quechua colexifies 'moth,' Fijian *kabu* also means 'to sow or scatter small seeds,' and for Bwe Karen *(dɛ)θochí*, perhaps compare *θochí* 'gallbladder.' In Bislama, the meaning of *sno* (< Engl. *snow*) has been extended to also mean 'fog.' Hawaiian *'ohu* colexifies "adorned with a leis," and *'ehu* inter alia also 'faint, difficult to see' and 'pollen.' Finally, Tetun has *rai-ahu* 'land-lime.'

26. The Forest

Representation: 84%
Motivated: 42.1%
Thereof Analyzable: 10.9% Thereof Colexifying: 31.3%
Thereof by Contiguity: 21.2% Thereof by Similarity: 0%
Recurrent associated meanings: bush/grove/scrub, wood, tree, mountain/hill, place/area, lumber/timber, branch/stick, plant, weeds, wilderness, thicket, taiga, place to defecate, grass, interior

As one might expect, terms for 'forest' (or 'jungle') are most frequently lexically associated with terms for 'tree' and/or 'wood,' either by colexification or morphological (semi-)analyzability (see Buck 1949 for evidence from Indo-European, among the languages where this association is attested diachronically is German). For 'tree' specifically, colexification is found in six languages: Ngambay, Khalkha, Upper Chehalis, Nez Perce, Yaqui, and Hawaiian (Ngambay, Yaqui, and Hawaiian, like Muna and Wayampi, colexify also 'plant,' and the Hawaiian term has also still other meanings). Alternatively, seemingly analyzable terms for 'forest' on the basis of 'tree' are found in eleven sampled languages, Anggor, Baruya, Kwoma, Comanche, San Mateo del Mar Huave, Santiago Mexquititlan Otomí, Arabela, Ancash Quechua, Fijian, in another Hawaiian term for the meaning, as well as in Tetun. In Baruya and Ancash Quechua, the relevant terms are of the lexical type, the second element bearing the general meaning of 'place' or 'area' (Baruya *yi'darya*, literally 'tree area' for 'clump of trees, forest' and Ancash Quechua *sacha marka* 'tree/plant area'), while in Anggor, *nɨmambe* 'forest, jungle' is analyzable as /nɨmɨ-ambe/ 'tree-in,' Kwoma *me kaba* "mature forest (including sago swamps) growing on relatively level terrain" might be a dvandva compound of *me* 'tree, wood' and *kaba* "type of long-bladed grass that grows prolifically in swamps and lagoons," in Comanche, the word for forest, *soo huuhpi,* is analyzable as 'many trees,' in San Mateo del Mar Huave, *tixiül* appears to be analyzable as 'AUG-tree,' Pipil (Cuisnahuat dialect) has *ku(:)htan*, analyzable as /kuh-tan/ 'tree/wood-locative/under,' and in Santiago Mexquititlan Otomí, 'forest' is *mbo-zaa* 'place.where.around-tree.' The Arabela term is formally the plural of that for 'tree,' Fijian has *vei-kau* 'COLL-tree/wood/stick,' and Tetun *ai-laran* 'tree-interior' (compare colexification of 'jungle' and 'interior' in Berik, presumably because this is where it is located). Hawaiian, moreover, has *ulu lā'au* 'grove tree/forest' (there is a further rare semianalyzable term featuring *ulu* which also means 'brush, undergrowth'), and there is another term that can mean either 'trees' or 'forest' (note that for Arabela and Cashinahua, the respective terms are also glossed as 'trees'). A Sora term for 'forest' is *ə'ra:'ja:ŋən*, containing *ə'ra:-* 'wood, wooden' and perhaps *'ja:ŋ-* 'bone, hard dry stalk.' In addition, there is a relatively large number of languages with semianalyzable terms for the meaning 'forest,' one of the constituents clearly being words for 'tree,' namely Kaluli, Xicotepec de Juárez Totonac, Bororo, Embera, and Sáliba.

Colexification with 'wood' is found in fifteen languages in the sample: Buli, Ngambay, Dongolese Nubian, Badaga, Basque, Chukchi, Greek, Khalkha, Laz, Welsh, Haida, Miskito, Hawaiian, Tetun, as well as Lesser Antillean Creole French (where the relevant

term also means 'arm'). This situation may also obtain in Wayampi, though here the metalanguage gloss 'bois' itself colexifies 'forest' and 'wood,' thus making the case unclear. Cubeo has *jocu-burumu* 'wood-vegetation.that.is.not.cut.down,' Fijian, as mentioned above, *vei-kau* 'COLL-tree/wood/stick,' and there are semianalyzable terms with an identifiable constituent meaning 'wood' in Sora, Xicotepec de Juárez Totonac and Maxakalí. Similarly, Khalkha, Cheyenne, Nez Perce, Pawnee, Hawaiian, and Lesser Antillean Creole French (also) colexify 'lumber, timber' more specifically.

As also discussed in § 6.3. dealing with environmental factors and their influence on patterns of colexification, eight languages colexify 'forest' and 'mountain' or 'hill' (see also Buck 1949: 46-47 for evidence from Indo-European): these are Kwoma, Laz, Nivkh, Sora (also colexifying 'clearing on hill'), Ineseño Chumash, Aguaruna, Huambisa, and Yay. Moreover, Kwoma colexifies 'forest' also with 'netbag' and 'womb,' Basque also with 'wooded hill,' and Khalka also with "mountain ridge, range, or plateau," alongside 'taiga,' the latter association being shared with Kolyma Yukaghir (note also that the potential cases of Xicotepec de Juárez Totonac, Copainalá Zoque, Aguaruna, Arabela, Aymara, Bora, Cashinahua, Cavineña, Cayapa, Chayahuita, Cubeo, Guaraní, Huambisa, Lengua, Sáliba, Tsafiki, and Yanomámi, where the relevant terms are (also) glossed as 'monte' and/or 'selva,' which are capable of referring to a 'mountain' but more generally denote all kinds of non-cultivated territory, and are hence ignored here when assessing colexification).

Quite a number of languages colexify 'forest' with meanings such as 'bush, bushland,' 'grove,' or 'scrub, undergrowth, underbrush.' These are: Bakueri, Buli (also colexifying 'savannah'), Efik, Hausa, Ngambay, Buin, Burarra, Gurindji, Kwoma, Nunggubuyu, Sko, Northeastern Tasmanian, Badaga, Blackfoot, Cheyenne, Haida, Ineseño Chumash (also colexifying 'chaparral,' a type of vegetation in California), Nez Perce, Pawnee (by a semianalyzable term containing an element meaning 'place'), Chayahuita, Miskito (also colexifying 'opening, hole' as well as 'rectum, anus'), Rama, Wichí, Hawaiian, Lenakel, and Samoan; Kanuri *kàráà sálám* is literally 'bush black.' Moreover, Oneida *kélhite?* is analyzable as /ke-lh-Nit-e?/ 'ANOMALOUS.PREFIX-woods-be.in-STAT.' Chickasaw *abokkolanonka'* is analyzable as /abokkoli'-anonka'/ 'thicket-inside,' while Khalkha and Hawaiian colexify 'forest' and 'thicket' directly.

Three Austronesian languages, Muna, Hawaiian, and Samoan, colexify 'forest' and 'weeds,' Muna and Wayampi also 'grass,' and Samoan 'tall grass' specifically. Furthermore, Ngambay colexifies 'branch,' and Khalkha, Nez Perce, Bororo, and Hawaiian colexify 'forest' with 'stick' (Nez Perce with 'small forest' more specifically), while in Carrier, *tcentherh* is analyzable as /tetcen-therh/ 'stick/wood-in.' While for Ngambay, Khalkha, and Nez Perce, it should be noted that the relevant terms also colexify 'tree' and the association with 'branch' may thus be not immediate, this is not the case for Bororo. Ngambay *kake* in fact has a very broad semantic range and may also refer to a 'rod,' 'stick,' as well as 'place' generally, a pattern shared with Jarawara, where the relevant term also means 'thing' and 'time' (see § 6.4.3.15.), and Kaingang. Similarly, Yanomámi colexifies 'forest' with 'region, area, territory' as well as 'city' and 'country,' and note also that the semianalyzable Lengua term mentioned above contains an element meaning 'place.' In Badaga, *ka:ḍu*, meaning 'forest, jungle woods,' but also 'field' and 'wasteland' among other meanings, is also used

as a term for the "place where one defecates," and in Lenakel, *nɨkinhamra* 'forest, bush' is also glossed as "[t]he place in the bush where one goes to urinate or defecate ..." Finally, relevant terms in Koyraboro Senni, Badaga, and Hawaiian are also glossed as 'wilderness.'

Other associations for this meaning and semantic nuances conveyed by the respective terms are manifold. Buli *sagi* also can refer to a far-away country inter alia, Hausa *kurmi* also denotes a "wooded ravine" and is an "expression of wonderment," and Ngambay *kɔr* also means 'coal' inter alia. Buin *maika* also is the name of two Buin villages, the Burarra term *gumurra* is based on the root *murra* meaning 'clustered.' Kwoma colexifies "dense undergrowth" among other meanings. The Kyaka term *imwua* can also refer to the ridge-pole of a house that is cut from a very tall tree, and Muna has *katugha ~ kamothuga*, derived from *tugha* 'hard, ripe' and *karumbu ~ karombu ~ kahombu*, perhaps related to *rumbu* "to stink (as of old wounds)." Another term colexifies 'forest, jungle' with "part of a field not yet cleared (of weeds)," as well as to "separate off part (of a field) to be worked later." Sko *hángpeng* apparently contains *peng*, which can either refer to a tree species or mean 'clear, cleared,' and Yir Yoront *larrkurrq* contains *larr* which means 'land, site' and other related things. Basque *baso* also means 'vase, glass' and 'vessel;' these senses are presumably due to borrowing from Romance. The Khalkha term *cabdaɣul* is derived from *cabda-* 'split,' using the agent nominalizer *-ɣul* and also means 'grove' as well as 'domino,' while *oi* also means 'birthday, anniversary,' and 'mind, intellect, memory' inter alia. Upper Chehalis *yámaćumš* appears to be analyzable as /yámac=umš/ 'douglas.fir=place,' while Nez Perce colexifies 'forest' with 'tree trunk.' The Wintu term *kel* can also be used in an adjectival sense, then meaning 'long, far.' Highland Chontal colexifies 'forest' with "wide open space outside the village, the woods," and more specifically also 'airstrip' and 'basketball court.' Copainalá Zoque *pecatza'ma* might be analyzable as /peca-tzaman/ 'old-amate.tree.' Guaraní *ka'aguy* appears to contain *ka'a* 'mate herb,' and Wayampi *kaʔa* also means 'grass.' Hawaiian *moku* can also refer to a 'district,' an 'island,' a 'fragment of something,' as well as, when used verbally, mean 'to be cut, broken' inter alia, while Rotuman *vao* also denotes a 'net, fishing net.' Sedang *chíu* also means 'burned' inter alia, *príu* also 'hailstone, ice,' *să* also "bow of kơtro bow trap," and Yay *doŋ*[1] also means 'to pickle.'

27. The Gold

Representation: 63%
Motivated: 34.2%
Thereof Analyzable: 20.3% Thereof Colexifying: 14%
Thereof by Contiguity: 17.7% Thereof by Similarity: 7.9%
Recurrent associated meanings: yellow, metal, money, silver, red, gold ornament, diamond in cards, jewel, iron

Motivated terms for 'gold' have a clearly recognizable areal distribution. They are rare in the Old World and common in the New World, in particular North America, but are sometimes also found in South America and Eastern Eurasia, including Southeast Asia. A frequent pattern, found in ten languages (Japanese, Khalkha, Cheyenne, Chickasaw, Nez Perce, Oneida, Pawnee, Bororo, Maxakalí, and Rama), is to have complex terms for 'gold'

where one of the constituents is a word for 'metal,' while in Nez Perce, Mandarin and Tetun, colexification of 'gold' with 'metal' is encountered. In analyzable items, the semantics of the second element is subject to some cross-linguistic variation. In Japanese, a term for 'gold' is *ō-gon* 'yellow-metal' (this structure is also found in Chickasaw, Kiliwa, Pawnee, where the relevant term colexifies 'brass,' Bororo, where the relevant term denotes any type of yellow metal generally, and Rama; in addition, note that Nez Perce has the optional complex term *maqsmáqs kícuy*, where *maqsmáqs* is 'yellow' and *kícuy* is 'metal, money, dollar, gold.' However, both constituents of the Nez Perce term can be used alone to refer to the concept 'gold'). In Cheyenne, 'gold' is called *véhone-ma'kaata* 'chief-metal;' lexicographers note that this denomination may have come into being due to "the Indian chief head which has appeared on some old gold coins," and note that this term may refer to gold coins specifically rather than the metal generally. Khalkha *erkim temyr* is analyzable as 'supreme iron/metal,' Maxakalí *pipkup yãnãm* as 'metal.object light/lamp,' Tetun *murak-mean* redundantly as 'metal/gold-red/golden,' Oneida *ohwistanolú* as /o-hwist-nolu-ʔ/ 'NOUN.PREFIX-metal-be.expensive-STAT,' and Pawnee also has the term *paapicistaariksisuʔ*, which is analyzable as /paapicis-raariksis-uʔ/ 'metal/money-authentic-NOM.' Both the Oneida and Pawnee terms in fact also mean 'silver,' and this is also the case in the Santo Domingo de Guzmán dialect of Pipil. In Lakhota and Malagasy, the relevant terms are analyzable (*mázaskázi* /mázaská-zí/ 'silver-yellow' and *volamèna* /vòla-mèna/ 'silver-red'). In fact, associations with the color terms for either 'yellow' or 'red' are also commonly found and that in combination with elements of varied semantics, as already seen in some of the above examples; this is also common in Indo-European (Buck 1949: 610). Alongside the already explicitly mentioned cases of Japanese, Chickasaw, Pawnee, Bororo, and Rama with terms of the 'yellow-metal'-type, as well as Lakhota and Nez Perce, an association with the color 'yellow' is found in other languages. Most often, the second element in terms of the analyzable lexical type is 'money,' as in Buli *ligmoaning* /ligra-moaning/ 'yellow/red-money.' Terms with constituents meaning 'money' and 'yellow' are also found in Kiliwa, Chayahuita, Miskito, Pawnee, and Rama (for the Nez Perce and Pawnee cases, compare discussion above). Terms with other or more specific patterns include Khoekhoe's *ǃhuni-lurib* 'yellow-iron,' Tuscarora's *tikači?tkwáhnayę?*, which is analyzable as /ti-ka-či?tkwahn-yęT/ 'PARTITIVE-3SG.INDEF.AGENT-yellow-lay,' and Guaraní's *ita-ju* 'rock-yellow.' The association with the color 'yellow' is by colexification in Sahu, Quileute, Hani, and Vietnamese (it cannot be excluded that in some of these cases 'gold' is meant as a color term rather than the name of the metal). Alongside the languages already mentioned, the association with 'red' is also found in Upper Chehalis (*čsčíq t tá·la* 'red INDEF.ART money,' *tá·la* being a loan from English *dollar*), and Rotuman *monē mi'a* 'money red.' As these examples make clear, a frequent constituent for terms for the 'gold,' by functional contiguity, is 'money.' This is found in a total of eleven sampled languages, most of which have already been discussed: these are Buli, Biloxi, Upper Chehalis, Comanche, Nez Perce, the Santo Domingo de Guzmán dialect of Pipil (colexifying also 'lead'), Pawnee, Miskito, Rama, Wayampi, and Rotuman (in addition, the Chayahuita term is semianalyzable and the recognizable constituent is a term for 'money'). The association is by colexification in Comanche, Nez Perce, Pipil, Pawnee, and Wayampi. In line with this, Oneida *ohwistanolú*,

which also means 'silver,' is analyzable as /o-hwist-nolu-ʔ/ 'NOUN.PREFIX-metal-be.expensive-STAT,' and Ancash Quechua colexifies 'gold' with 'very valuable, precious.'

Finally, three sampled languages, Hausa, Badaga, and Bororo, by provenience or configurational contiguity, colexify 'gold' with 'gold ornament' or particular ornaments made of gold (Bororo also with a type of labret), and two further languages, Carib and Sedang, colexify 'gold' and 'jewel.' Wappo *ʔóˑroʔ* and San Lucas Quiaviní Zapotec *oor* (both < Span. *oro*) also mean 'diamond in cards' (San Lucas Quiaviní Zapotec *oor* also means 'time, hour' due to phonological collapse of Span. *oro* and *hora*). Similarly, Basque colexifies 'gold' with 'gold suit' (in the Spanish Baraja deck of cards).

Further motivated and possibly motivated terms for 'gold' include: Efik *kut* inter alia also means 'to see, to feel,' Ngambay *tùngu* is also the name of a brown fruit, and Muna *bulawa* may be related to *bula* 'albino, white skin.' Sora colexifies 'gold' with 'wealth.' Comache *puhihwi* is analyzable as /puhi-ekahwi/ 'leaf-shiny;' since this term also colexifies the meaning 'money,' a possibility is that this term originally referred to a banknote and was only later extended to 'gold.' Wintu *yoˑqas* contains *yOq* 'wash, pan for gold, gold dust,' and Carib *ikamipo* is derived from *kami* 'to glisten' by the circumfix nominalizer *i-...-po*. Bislama *gol* also means 'goal' (< Engl. *gold* and *goal*), and Bwe Karen *thé* also 'bear' inter alia. Hawaiian colexifies 'gold' with 'plain,' 'pasture' and other meanings (the 'gold'-reading of *kula,* the relevant term, is due to secondary English influence). Likewise due to the collapse of an English loan with native lexical material, *koro* also means 'fortress' inter alia. Finally, White Hmong *kub* also means 'horn' and 'to burn,' while Yay colexifies 'gold' with "loved or treasured person" and 'needle.'

28. The Grass

Representation: 88%
Motivated: 37.2%
Thereof Analyzable: 2.9% Thereof Colexifying: 34.2%
Thereof by Contiguity: 19.0% Thereof by Similarity: 0.8%
Recurrent associated meanings: hay/straw, meadow/pasture/lawn, plant, vegetable, green, leaf/foliage, (garden) rubbish, shrub, undergrowth, tobacco, fodder, forest

Lexico-semantic ties between 'grass' (or 'weeds, herb' – these are not distinguished for present purposes, and neither is 'grass' collectively as opposed to individualized 'blade of grass') and other concepts are typically realized by colexification, where the underlying semantic relation is one of contiguity. Globally, the most frequently found pattern (in 22 languages) is that languages feature a single term for 'grass' and 'hay' and/or 'straw.' This is particularly frequent in North America, but also figures in a significant number of languages in other parts of the world, namely in Buli, Ngambay, Abzakh Adyghe, Badaga, Bezhta, Khalkha, Kildin Saami, Upper Chehalis, Cheyenne, Chickasaw, Highland Chontal, Itzaj, Kashaya, Kiowa, Lakhota, Nez Perce, Nuuchahnulth, Oneida, Pipil, Tuscarora, Yaqui, and Mandarin (as well as in Indo-European languages, Buck 1949: 519-520). Yahi and Central Yana *maadu* is also glossed as 'wild hay.' Turner and Brown (2004) interpret the presence of this pattern in Salishan and other languages of North America as a process of lexi-

cal acculturation: since 'hay' as a foodstuff for domestic animals was unknown in pre-contact times, there was no distinct term for it, and the native terms for 'grass' experienced semantic extension to also cover this aspect.

In fourteen further languages, Buli, Rendille, Kwoma, Ngaanyatjarra, Toaripi, Badaga, Greek, Khalkha, Welsh, San Mateo del Mar Huave (colexifying also 'firewood'), Arabela, Chayahuita, Kaingang, and Tetun, terms for 'grass' are found that at the same time also may be used to refer to a conjunction of grass, that is, a 'meadow,' a 'pasture,' a 'lawn' or a grass-covered area in general (in Buli, this meaning is associated with the plural form of the term). This is a fine instance of configurational contiguity (also attested diachronically in Indo-European according to Buck 1949: 519).

Just a little less frequent, occurring in thirteen languages, namely Muna, Abzakh Adyghe, Badaga, Bezhta, Ket, Kildin Saami, Upper Chehalis (by a semianalyzable term containing an element meaning 'grow'), Cheyenne, Ineseño Chumash, Embera (colexifying also 'thicket,' 'vegetation,' as well as 'mount, mountain;' the meanings are associated with different genders), Guaraní, Maxakalí, and Wayampi is colexification of 'grass' with '(green) plant' (sometimes also 'bush' and other similar meanings), that is, colexification of two adjacent ranks in the ethnobiological taxonomy (the relevant Muna and Wayampi terms also mean 'forest,' while similarly Rotuman colexifies 'grass' with 'plants and bushes growing wild' inter alia). Embera and Fijian also colexify 'shrub' more specifically, and Kaluli and Rotuman also 'undergrowth.'

In four sampled languages, Rendille, Ngaanyatjarra, Khalkha, and Tuscarora, the term for 'grass' is associated lexically with the color term for 'green' (it appears to be quite frequent for color terms for 'green' to be lexically related to terms for 'grass,' see e.g. Callaghan 1979 on some Miwokan languages, Proulx 1988 on Algic languages, and Buck 1949: 520 for diachronic evidence from Indo-European). In Tuscarora, there is also a complex term for 'grass' on the basis of the root *–her-* 'green, grass,' namely *uherúhkweh*, analyzable as /u-her-(a)hkw-eh/ 'NOUN.PREFIX-green/grass-INSTRUMENTAL-NOUN.SUFFIX.' This term has a wide semantic range, which includes alongside 'grass' also "ground, cover, hay, reed, rush" as well as 'weed.' Alongside these languages, there is a semianalyzable term for 'green grass' found in Welsh which appears to contain *glas* 'green.'

Also in four languages, namely Efik, Swahili, Tasmanian (Middle-Eastern, Southeastern, and Western), and Xicotepec de Juárez Totonac, and here exclusively by colexification, an association with the meaning 'leaf' and/or 'foliage' is found (in Swahili also with 'needle'). Other recurrent associations by colexification, some of which are particularly frequent in certain regions of the Old World include that with '(garden) rubbish' (Baruya, Buin, Sora, Bora, in which latter also "cumbering things" is colexified), with 'tobacco' (Efik and Basque, which latter also colexifies 'brow, forehead'), with 'fodder' (Badaga, Sora; an association very common in Indo-European, Buck 1949: 519-520), and with 'vegetable' (Efik, Ket, Khalkha, Kildin Saami, Guaraní, and Fijian, compare also Greek *chortári* 'grass' and *chortarikó* 'vegetable' and the colexification with 'lettuce' in Nez Perce).

Associations found only in one particular language only are many. Efik *i'köñ* also means 'bud,' and Hausa *haki* also means 'wages' and may refer to "panting; gasping, breathlessness." *Ciyawa*, another Hausa term, is also used to refer to the first signs of vari-

ous diseases. Khoekhoe colexifies 'sheath of grass,' and Noni *gayi* is also used to refer to so-called "elephant grasses" specifically. Rendille *roób* can also be used to characterize the fertility of vegetation, Anggor *wohɨ* also means 'to weed,' and Kaluli *kis* also may refer to 'small weeds in garden.' Nunggubuyu *maḏa* also refers to a fish species called 'long tom' which "swims with long, pointed mouth out of water, perhaps resembling a blade of grass." Rotokas *rauritave kou* and *isiso kou* may contain *kou,* a classifier for 'heaps' according to Robinson (2011), and *isiso* is "steaming, giving off steam," "steam something, heat with steam, scald." Tasmanian terms in all varieties except the Western also denote 'seaweed,' while Yir Yoront *warrq* exhibits a typical Australian actual/potential polysemy, meaning both 'grass' and 'grass or bush fire.' Abzakh Adyghe *wɔc* is also can also mean 'medicine' (presumably originally made from herbs, i.e. by metonymy) and then, by further extension also 'chemical medicine' and 'chemical product' in general. Badaga *aṇe ~ haṇe* can also mean 'grazing' and 'swamp, wetland,' as well as "upper part of a ridge" and 'water channel,' Basque *belar* colexifies 'forehead,' and Khalkha 'dish' of food. Kildin Saami *rāss'* also means flower,' Upper Chehalis *smáqʷmumš* is analyzable as /s-máqʷm=umš/ 'CONTINUATIVE-prairie=place,' and Nuuchahnulth *ʕaqmapt* also denotes the 'Brome grass' specifically. For Kiowa *sǫ́ụ́dɑ́* 'plant or tuft of grass,' compare *sǫ́ụ́-dɑ́* 'grind; brush hair,' and for San Lucas Quiaviní Zapotec *gyihzhya'ah*, compare *gyihah* 'rock, stone.' Tuscarora *unę́hsakwt katerʔahθę́·tih* must be analyzable (*unę́hsakwt* may mean 'by the house'), but the constituents cannot be identified with any certainty. Wintu *ċaru·q* 'grass, greens' also denotes 'edible clover' specifically. Arabela *jiya-socua* is one of the few cases of a complex term for the concept, literally meaning 'land-CLASS.ROUND.TIMBER.' Guaraní *ka'a* also denotes a "type of Paraguayan tea" and 'vegetation' generally. Bislama colexifies 'grass' with 'feather,' 'pubic hair,' 'fern' and 'mould,' Fijian with 'to pierce' inter alia, and Bwe Karen with 'to command' and other meanings. Hani *jahhaq* might be related to *haq*, meaning inter alia 'bitter,' Hawaiian *mauʻu* is also the name for a 'kava strainer' and 'strand of pandanus plaiting' used in making hats, while Mandarin colexifies 'grass' with "disorderly, negligent." Finally, Lesser Antillean Creole French *zeb* can also mean 'black magic.'

29. The Headland

Representation: 26%
Motivated: 47%
Thereof Analyzable: 30.8% Thereof Polysemous: 18.8%
Thereof by Contiguity: 10.3% Thereof by Similarity: 29.5%
Recurrent associated meanings: nose, point/tip, head/forehead, end/to end, corner, sprout

'Headland' (or 'cape, promontory') is one of the meanings predominantly expressed by metaphor-driven analyzable terms. In fact, the most frequent association is a fine example of a body-part metaphor. This is the association with 'nose,' clear cases of which are found in six of the 39 languages for which data are available. This association is realized by analyzable terms in two languages, Basque and Vietnamese (*lur-mutur* 'earth-snout/nose' and *mũi đất* 'nose land' respectively), and by colexification in Kolyma Yukaghir, Haida (in

which the relevant term can also inter alia mean 'end' in general, for which compare colexification with 'to end, be finished' as well as "ultimate, final, last" and "hair, lady's locks, tuft of hair, knot" in Badaga and Welsh *pen-rhyn* 'end/head/top/mouth-hill' and *pen-tir* 'end/head/top/mouth-land'), Fijian (colexifying also 'mountain peak' inter alia), and Rotuman (which also has optional complex terms). In addition, the Ineseño Chumash and Great Andamanese terms are semianalyzable, and one of the identifiable constituents is the respective word for 'nose.' In four languages, another body-part metaphor is found, the source concept this time being the 'head' or 'forehead,' for which compare the etymology of Engl. *cape*, ultimately related to Latin *caput* 'head' (Oxford English Dictionary). The relevant languages are Yoruba (*iyọrí ilẹ̀ ògógoro*, analyzable as /i-ọrí ilẹ̀ ògógoro/ 'NMLZ-to.sprout.or.appear head'), Welsh (*pen-rhyn* 'end/head/top/mouth-hill' and *pen-tir* 'end/head/top/mouth-land'), Lavukaleve, and Hawaiian, where the association is by colexification (but in Lavukaleve it is more precisely 'forehead' which is colexified, and in Hawaiian also 'brow' and 'wisdom' inter alia are colexified). In addition, Swahili *rasi* assumes the meaning 'head' in poetic languages (this association is already present in Arabic, from which it is borrowed), and the case of Toaripi *harihari* is somewhat similar to this pattern, as this term also denotes the 'temples' (as well as 'river bend'); the same goes for Efik, where *isü* in *isü iköt* can, alongside 'tip of tongue, tip of finger' and other meanings in a variety of contexts, also mean 'face' (*iköt* is 'forest'). Two sampled languages, Tuscarora and Mandarin, have complex terms for the meaning 'headland' in which 'corner' acts as a source concept, for instance, the Mandarin term is *hai3-jiao3* 'sea-corner/horn' (the Tuscarora term colexifies 'cave' as well as 'nook'). Another case of shape-based similarity with a body-part is found in Takia *damo-*, which also means 'shin bone.' Four languages' terms betray associations not characterized by metaphor, in that the source concept is 'point' or 'tip': these are Efik (for data see above), Buin (*tiutiuna*, reduplicated from *tiuna* 'point, peak of mountain, junction of rivers'), Ket (*solgup* /saʔl-kūb/ 'sharp-point' for a "piece of land jutting out into river"), Kashaya (*ʔama· pʰišušaʔ*, containing *ʔama·* 'land' and *-ʔšuš-* 'be pointed'), and Rotuman, where in fact 'nose' and 'point, tip' are both colexified with 'headland' alongside other meanings. Moreover, in Hawaiian one term for 'headland,' *ʻoiʻoina*, is a nominalization of *ʻoiʻoi* 'pointed, protrude, sharp,' and in Aymara, *moqo* may also refer to a 'protuberance' of any kind. The Tetun term for 'headland' is *rai-dilan* 'land-bud/sprout/shoot,' for the similar conceptualization in Yoruba see above.

Further patterns in the available data are: Efik *ukabare* is derived via the nominalizing prefix *u-* from *kabare* 'to turn' and also means "a turning around; varying, variation" as well as a 'turning point' and 'turn' generally. Buin colexifies 'cape' with 'peninsula,' and Rotokas *kipekipea* also means 'point at base of ridge' and 'spine of lizard.' Sko *pong* can also function as a verb meaning 'to be closed' and 'to blow at fire,' but this may be due to homonomy rather than a genuine case of lexical motivation. Chukchi colexifies 'prow of boat.' Nivkh *knyk* also means 'rock, cliff,' and Nuuchahnulth *ʔapquuʔa* is also a toponym denoting a specific point of land. Malagasy *tànjona* can also mean 'aim' or 'objective' and *tsìraka* also denotes a "sandy seashore," while Samoan *tolotolo* is reduplicated from *tolo*, which has a large amount of different meanings, among them 'to crawl,' 'to pull,' 'to mix,'

and also 'yam tuber,' which is perhaps the most likely reduplication base because of similarity in shape.

30. The Honey

Representation: 77%
Motivated: 53.2%
Thereof Analyzable: 27.7% Thereof Polysemous: 25.4%
Thereof by Contiguity: 31.7% Thereof by Similarity: 14.1%
Recurrent associated meanings: bee, liquid/sap/juice, sweets/sugar, beehive, syrup, fat/grease, resin, honeycomb, faeces, nectar, wax, jam, gum, oil, egg

'Honey' is an interesting concept cross-linguistically in that in some languages, the same term is actually used to denote both the product 'honey' as well as its producer, that is, the 'bee.' This pattern is found in 17 of the sampled languages, namely Hausa, (denoting "honey together with the comb"), Ngambay, Yoruba, Burarra, Ngaanyatjarra, Nunggubuyu, Yir Yoront (by a number of semianalyzable terms denoting different types of bee species and the honey they produce, all containing *may* 'vegetable food,' one of them also containing *ponh* 'hornet' and another one containing *kathn* 'yamstick'), Chickasaw, San Mateo del Mar Huave, Xicotepec de Juárez Totonac, Cavineña, Kaingang, Lengua, Miskito, Wichí, Yanomámi, and Hawaiian, and thus in all areas of the world except Eurasia, including Southeast Asia (compare the absence of the pattern in Indo-European, Buck 1949: 383-384); moreover, Bislama *honet* 'wasp, hornet' rarely assumes the meaning 'honey.' There are also languages with a variant of this pattern, namely languages in which the meanings 'beehive' and 'honey' are colexified. These are Gurindji, Itzaj, and Yanomámi, while in Bora and Chayahuita, the association with 'beehive' is realized by complex terms involving classifiers (*íímúhojpácyo* /íímúho-hpácyo/ 'beehive-CL.liquid' and *nino-i'* 'beehive-CLASS.LIQUID' respectively; Bora *íímúho* 'beehive' is in fact itself analyzable as 'sweetness-CL.oblong,' and the language has the alternative term *iimúbá arááve-jpácyo* /iimúbá arááve-hpácyo/ 'sugar.cane condense-CL.liquid,' with *iimú-bá* 'sugar cane' analyzable as 'sweetness-CL.3d'). Moreover, Sora has *ədaŋ'da:n ~ adaŋ'da:n*, presumably containing elements meaning 'beehive' and 'water, liquid,' and Cubeo has *mumicoro*, consisting of the root *mumi* associated with bees and the classifier *-coro* for liquid states, and in Yanomámi all three semantically related meanings – 'bee,' 'honey,' and 'hive' – are expressed by the same term *puu*. The same situation obtains in Nunggubuyu.

However, even more frequent than colexification with 'bee' are complex terms in which this meaning serves as a contiguity anchor. Within this class of terms, several recurrent patterns are encountered. By far the most common cross-linguistic tendency, encountered in the form of fully analyzable terms in ten sampled languages, Kyaka, Meyah, Rotokas (here, 'bee' and 'honeycomb' are colexified, while in Nunggubuyu and Kiliwa, 'honey' and 'honeycomb' are, in Kiliwa by the analyzable term *mi?=yaw-y* 'light=larvae-ATT'), Toaripi, Pawnee, Yaqui, Maxakalí, Piro, Rama, and Tetun, is to have terms for 'honey' consisting of the word for 'bee' (or rarely 'wasp') and 'water,' 'liquid,' 'sap,' or 'juice.' Note also that Hausa, which is of the colexifying type, also features the

compound *ruwa-n zuma* 'water/liquid-GEN honey/bee,' denoting "honey apart from the comb" specifically; the situation in Miskito and Hawaiian is comparable. However, notably, even this association is not exclusively realized by morphologically complex terms. In Nunggubuyu, *-(w1)aṛgayag* (which may be related diachronically to *-aṛ-* 'water') also means 'juice,' Basque *ezti* means both 'sap, juice' and 'honey' (alongside 'graft, grafing' and 'vaccine' inter alia), and in Huambisa, *yumiri* is used with reference to both 'liquid' in general as well as 'soup,' 'juice,' and 'honey' specifically. There are redundant terms with this structure on the basis of the term colexifying 'bee' and 'honey' in Yanomámi and Hawaiian. A semianalyzable term where the identifiable constituent is 'water' is found in Berik.

In four languages (Efik, Mbum, Biloxi and Carrier), an association by complex terms with 'fat' or 'grease' is found; for instance, in Carrier, 'honey' is *tṣihna-rĕ·* 'wild.bee-grease.' In addition, in Nunggubuyu, the association is realized by colexification, and Ngambay has a complex term of this kind which is however redundant due to colexification of 'bee' and 'honey.' In two languages, Efik and Kwoma, an association with 'oil' is found, for instance in Kwoma *heemi kwar* 'bee oil' (*kwar* is more precisely "the name of two related types of tree with very oily wood;" the fact that Efik figures in both categories is because the consituent element in fact can refer to either 'fat' or 'oil'). A semantic pattern either realized by analyzable terms or by colexification is the association with the meanings 'sweet(s)' or 'sugar.' Next to the somewhat different case of Bora already discussed above, it occurs by means of morphologically complex terms with the other constituent being 'bee' in Upper Chehalis, Comanche, and Hani (for instance, in the latter language one of the terms for 'honey' is *biaq-qul* 'bee-sweet;' in Comanche, ~~*unu*~~ *bihnaa* is analyzable as /~~un~~ú? pihnáa?/ 'insect sugar' more specifically), and by colexification in Muna, Basque, San Mateo del Mar Huave, Kiowa (colexifiying also 'panocha'), Lake Miwok, Yaqui, Aymara, Guaraní, Hupda, and Imbabura Quechua. Xicotepec de Juárez Totonac colexifies 'honey' with 'sugar cane conserve,' Pipil (Cuisnahuat dialect) has *-neka:yut,* containing *nek* 'candy, sweets' and *a:yu* 'liquid, juice.' In addition, the semianalyzable Abzakh Adyghe term *ṡ°ew* contains *ṡ°e* 'sweet, sugar,' and redundant terms with this structure on the basis of a simplex colexifiying 'bee' and 'honey' directly are encountered in Chickasaw and Miskito. A similar pattern predominantly found in North America (Upper Chehalis, Cheyenne, Itzaj, Nez Perce, Oneida), but also in White Hmong in Southeast Asia is that with 'syrup.' For instance, in Upper Chehalis, *cíčsməlàs* is analyzable as /číčs-məlás/ 'bee-syrup,' *məlás* being a loanword from French (< *mélasse*) that entered the language via Chinook Jargon. The association is by colexification in Itzaj, Nez Perce and Oneida.

Haida *sraal narii* is analyzable as /sraal nara/ 'bee faeces.' Other languages with terms with such structure are Blackfoot and Fijian. In one sampled language, Buin, the same term is used for 'urine' and 'honey' (the relevant term is also a toponym for a river and a mountain), a pattern which one might have expected to be more frequent given that, alongside the tertium comparationis available for both 'faeces' and 'urine,' namely that all three are excretions of a living being, here also the fact that both 'honey' and 'urine' are fluid is available. In Dadibi and Bislama, an association by morphologically complex terms with 'egg' is found; in Dadibi 'egg' is colexified with 'nut' and 'round object generally,' and in Bislama, the term denotes especially honey "taken directly from the

hive," while otherwise *hani* (< Engl. *honey*) is also used. Wappo is unique in the sample in that in this language, 'honey' is called *tuš-humév* 'bee-tear.' There is a semianalyzable term where the identifiable constituent denotes a certain bee species in Piro, and one with the generic term for 'bee' in Hani.

By colexification, three sampled languages associate the meanings 'honey' and 'resin.' These are Nez Perce, Oneida, and Bwe Karen (colexifying 'to be pure, clean,' inter alia), and Tuscarora has *ruʔtáhkę uθrę·weh*, consisting of *ruʔtáhkę* 'bee' and *uθrę·weh*, which can refer to 'resin,' but also 'cement,' 'glue,' 'gum,' 'jam, jelly,' 'molasses,' 'syrup,' 'tar,' and 'wax' ('jam' and 'honey' are colexified directly in Gurindji, and 'gum' and 'honey' are in Oneida; see above for the association with 'syrup'). As is clear from this list, this pattern appears to be particularly frequent in the Americas. Buli colexifies 'honey' and 'wax' (and this may be the case in Nunggubuyu, too, but this is not sure), and Khalkha and Vietnamese share colexification with 'nectar.'

Further unique associations in the sample include the following: Koyraboro Seeni a*yuu,* a dialectal variant of *yuu*, also means 'manatee,' and for Ineseño Chumash *'aq'ika'š* compare *aq'ika'š* "to burn in the throat from sweetness, as honey does." Nunggubuyu *mala* also means 'navel' and '(clear) sky,' Ket colexifies 'honey' with 'copper,' and Nez Perce *te·mísquy*, alongside 'syrup' and 'honey,' also means 'sorghum' and denotes the "gelatinous sap of the tamarack tree." Bora, alongside a term featuring an association with 'beehive,' also features other terms formed with the classifier for liquids, *-hpácyo*, namely *ócóómého*j*pácyo,* where *ócóómého-* denotes the hive of a particular type of bee. The Wayampi term *ɛi ~ ɛy* also can refer to a religious 'honey dance' that the Wayampi perform. Manange *4k^{wh}e* also means 'song,' and Hawaiian *hone* (at least in the sense of 'honey' < Engl. *honey*) colexifies 'honey' with "sweet and soft" (said of music) and "sweetly appealing" (said of perfume) and other meanings. Furthermore, there are a number of terms in different languages in which the term for honey strikingly resembles semantically related words, but which resist morphological analysis. These include the following pairs: Highland Chontal *algujua* 'honey' and *galgujua* 'honey bee,' Wintu *hu·bi* 'honey of bumblebee' and *hu·bit* 'yellowjacket' (sharing the root *hu·b-*), Huambisa *yumiri* 'liquid, soup, juice, honey' and *yumin* 'sweet,' and finally Lengua *yohena* 'honey' and *yohan* 'bee.' Finally, it should be noted that several sampled languages have a more specific honey terminology, in which one general term stands alongside a range of more specific terms for different types of honey, and in Arabela, a general superordinate term is lacking altogether and only specific types of honey are lexically designated.

31. The Horizon

Representation: 28%
Motivated: 66.5%
Thereof Analyzable: 67% Thereof Colexifying: 1.9%
Thereof by Contiguity: 3.7% Thereof by Similarity: 54.9%
Recurrent associated meanings: sky, edge/border/fringe, end/finish, land/earth, meet/meeting place, cloud, basis, sea

As the above figures make clear, the 'horizon' is one of the concepts predominantly expressed by analyzable terms driven by metaphor. The most frequent lexico-semantic association is that with meanings such as 'edge,' 'border,' or 'fringe.' This is found by colexification in Buin (though the relevant term *kitai(na)* appears to be derived from *kita* 'split longitudinally, split with the grain') and by analyzable terms with 'sky' acting as contiguity anchor in Abzakh Adyghe (there is also an alternative version of the term where 'sky' is replaced by 'cloud'), Basque, Khalkha, Carrier, Haida (where the term can be literally translated as "(where) the edge of the sky hits the land or water"), Bororo (which also has the alternative term *barubaru*, reduplicated from *baru* 'sky' and colexifying 'type of spirit'), Kaingang, and Hani (where *aoqkaq milkaq* contains *aoq* 'sky' and *milkaq* 'border line;' there also is another term with similar structure featuring *zeil* 'line'). There is a derived term in Hawaiian which also denotes the "[c]ords or fine ropes threaded through marginal meshes of upper and lower edges of nets" and also means 'trickery, deceit, deceive, cheat,' and similarly, Greek *orízōn* is related to *orízō* 'to bound, delimit.'

A very similar conceptualization strategy where 'end' is the source concept is found with 'sky' as the contiguity anchor in Efik and Guaraní (*utït enyöñ* 'end sky' and *arapaha* /ára-paha/ 'sky-end' respectively). Similarly, Maxakalí has *hãmnõgnõy*, presumably containing *hãm* 'earth' and *nõg* 'finish off, use up.' In Buin, the association is by colexification; it is the only sampled language to behave in this way. In three languages, Upper Chehalis, Miskito, and Hawaiian, an association with verbs meaning 'to meet' or nouns meaning 'meeting-place' is found. For instance, Upper Chehalis *x̣ʷuqʷx̣ʷuqʷɬ́lwstn* contains *x̣ʷúqʷi-* 'to gather, come to be gathered' and the lexical affix *=tn* 'place.' In Hawaiian, this association is by colexification. Furthermore, in three sampled languages of Oceania, Kosarek Yale, Fijian, and Hawaiian, analyzable terms featuring an element meaning 'basis' are found (Fijian *vū-ni-lagi* 'basis-POSS-sky,' Hawaiian *kumu-lani* 'base-sky,' Kosarek Yale *imbubuak* and *soobubuak*, containing *im* 'sky' and *soo* 'earth, soil' respectively alongside *bubu* 'liver, carrying part, point of support'). Also in three sampled languages, one of the constituents of the term for 'horizon' is that for 'cloud.' One of the Abzakh Adyghe terms for the 'horizon' is, as alluded to above, *pšape* /pše-pe/ 'cloud-edge/tip,' in Biloxi, it is called *natcí˝-xwŭhí˝* 'cloud-low' (this term indeed also denotes a 'low cloud'), and in Itzaj *chäkil muyal* 'red cloud.' As has already become clear from the previously mentioned Kosarek Yale and Haida terms, in some languages a constituent of the word for the 'horizon' is a term with the meaning of 'land' or 'earth.' This is also found in Maxakalí and Miskito. Further terms where one constituent is 'sky' are Ngambay *gél-dàra* 'origin-sky,' Khalkha *tngri jin xormai* 'sky GEN lower.hem/foot.of.mountain,' Kiliwa *ʔ+maʔi=haa=p+ʔiw-m-uʔ* 'DN+sky=move/go=PASS+stand-THR-OBL,' a more palpable literal translation of which is "where the sky stops," Tuscarora *weyurẹhyaʔníhẹ* (built around the roots *-rẹhy-* 'sky' and *-ʔniha-* 'sprain'), Bororo *barubaru* (reduplication base: *baru* 'sky;' the complex term is also the name for a kind of spirit), Hani has *aoq-daoq aoq-zeil*, containing *aoq* 'sky, heaven' and *zeil* 'line,' Hawaiian *pōʻai-lani* 'circle-sky,' Lenakel *noua-nɨsii-neai/tehe* 'fruit-excrement-sky/sea' (for which compare Carib *palana lalɨ* 'sea floor' and Hawaiian *ʻili-kai* 'surface-sea,' which indeed also denotes the 'surface of the sea' and 'horizontal' more abstractly), and

Rotuman *lā ne lạgi* 'foot ART.PL sky.' There are semianalyzable terms in Central Yup'ik, Cubeo, and Guaraní.

Further associations include: Khoekhoe *mûs (di) lkhâu-s* is analyzable as 'eye (POSS) radiate-3SG.FEM' (Khoekhoe has also borrowed *horisonni* from English or German), Nuuchahnulth has a lexical suffix colexifying 'on the horizon' with 'far out at sea,' and Chayahuita *hui'tontarinso'* contains *hui'tonin-* 'paint with horizontal strokes' (for which compare colexifiation of 'horizontal' as well as 'surface of the sea' in Hawaiian), and there is another term, *aquë notëhuatëra,* containing *aquë* 'far.' Maxakalí *hãpkux-yã* may be analyzable as 'shoreline-fragment,' for Hani *caqqiq*, compare *caq* 'a mark' and *qiq* 'to lift up' (both possible constituents also have further meanings). Hawaiian, featuring a wealth of terms for the concept, also has *Kahiki moe* 'Tahiti prostrate,' as well as *ʻalihi lani*, which is redundantly analyzable as 'horizon sky' and colexifies "deceit, trickery" inter alia. Samoan *i tafatafa-ʻila-gi* is analyzable as 'PART side-spot-SUFFIX' and means 'on the horizon'[4]

32. The Horn

Representation: 86%
Motivated: 44.7%
Thereof Analyzable: 7.2% Thereof Colexifying: 37.5%
Thereof by Contiguity: 12.2% Thereof by Similarity: 24.3%
Recurrent associated meanings: antler, musical instrument, branch, tusk, antenna, point, ear, tooth, cockscomb, bone, post, tributary, clavus, nail, thorn, signal horn, corner, wood

Analyzable terms for the meaning 'horn' are relatively rare cross-linguistically, and where they occur, it is typically in areas without indigenous species of large horned animals, such as Australia and the Amazon region of South America. An association with 'ear' is found by complex terms in Jarawara (*warabi awe/warabi ewene* /narabi/narabo awe/ewene/ 'ear piece.of.wood,' for which compare Kaingang *nĩ-ka* 'flesh-tree/wood'), Wayampi (*ãmilãsĩ*, containing *nami* 'ear' and *ãsĩ* 'pointed'), and Hawaiian (*pepeiao-hao* 'ear-iron,' with *hao* also being the name of the horn of a goat itself). Similarly, Yir Yoront has *pin+ngon* 'ear+hornlike.process;' in addition, the association is realized by colexification in Burarra (where the relevant term may also refer to any "appendage that sticks out like an ear" as well as "witness to what was said"). Great Andamanese has *wôlo-tâ* 'adze-bone.' This term denotes the 'horn of cattle' specifically, and the translation of the example for its usage provided in the source makes particularly clear that it is a neologism for a newly encountered concept: "when we first saw cattle we called the horns (lit. things on their heads) *wôlo-tâ (da)*, i.e. adze(-like) bones." By colexification, the association is found in Mesoamerica, namely in Highland Chontal and Itzaj (in the latter language, *b'ak*, the term in question, also means 'employment;' the association with 'bone' is also noted for Kolyma

[4] The source for Yoruba also lists *ibiti o dabi ẹnipe ilẹ ati ọrun pade* 'place:REL 3SG appear like land and sky meet;' it is not mentioned in the above since Joseph Atoyebi (p.c.) informs that this is not a conventionalized part of the Yoruba lexicon.

Yukaghir according to an older source incorporated into the consulted source). In fact, the association with 'ear' and 'bone' is the only recurrent one typically realized by morphological analyzability. Further analyzable terms include Central Yup'ik *ciru-neq* 'cover-area,' Bora *iyáábej-to* 'animal-CL.spine,' and Kapingamarangi *madaagoo* /mada-goo/ 'end-husking.stick/coccyx.'

Associations by colexification, in contrast, are much more varied cross-linguistically, although also here recurrent tendencies emerge. Basque, Chukchi, Nivkh, Kolyma Yukaghir, Carrier, Upper Chehalis, Cheyenne, Haida, Kiowa, Lake Miwok, Nez Perce, Quileute, Wappo, Wintu, Yuki, Central Yup'ik, Tehuelche, Bwe Karen, Hani, Hawaiian, and Sedang colexify 'horn' with 'antlers.' In five languages, Rotokas, Chickasaw, Ancash Quechua, Wayampi, and Tetun, the word for 'horn' is also used with the meaning 'antenna' or 'feeler' of an insect, or, according to the Ancash Quechua source, of a 'worm' in this language (compare also Sora *dereŋən* 'horn' and *dereŋbudən* 'antenna of insect'), while six languages, Hausa, Ngambay, Basque, Kildin Saami, Abipón, and Toba, have one term covering both 'horn' and 'branch' or similar meanings (Basque also colexifies 'tributary' - a pattern also found in Toba- 'lineage,' and 'bedpost, chairpost'). Furthermore, in three languages, Arabela, Ancash Quechua and Bislama, the same term is used for 'horn' and 'cockscomb' (in Bislama apparently also for 'comb'), and in two, Khalkha and Lesser Antillean Creole French, 'horn' has a semantic extension to 'clavus.' In two further sample languages, Wintu and Fijian, an association with 'nail' is present (in Fijian by the analyzable term *i vako* 'DERIV put.nail.through,' while in Wintu, the same term may also be used to refer to a 'hammer' and a 'drill'). Also exclusively by colexification, in three languages, Kyaka, Chukchi, and Samoan, 'tooth' and 'horn' are lexically associated (the Kyaka term also colexifies 'food, edible material' as well as "sharp, biting, erosive," and the Samoan term *nifo* is also used figuratively to refer to an 'enemy'). Similarly, Takia *fai* colexifies 'horn' with "upper canine teeth" specifically, and also with 'crocodile.' Kyaka (where the relevant term also has still other meanings) and Guaraní colexify 'horn' with 'thorn.' Similarly, Lake Miwok *kílli* 'horn, antler' seems to be related to *kíili* "to hook with the horns, thorn, stickers on weeds or berry bushes." In addition, Bora and Guaraní colexify 'horn' and 'point' (compare Wayampi's term *amilãsĩ* containing *ãsĩ* 'pointed,' and the similar connection with 'point, peak' in Irish mentioned by Buck 1949: 208), while Aymara and Basque colexify 'horn' and 'post' ('bedpost,' 'chairpost' more specifically in Basque). For six languages, Kanuri, Basque, Welsh, Chickasaw, Lesser Antillean Creole French, and Hawaiian, lexicographers note that the word for horn is also used for a musical instrument, as is the case in many European languages (Welsh *corn* also colexifies 'corn,' 'stethoscope,' and 'might'). Furthermore, in Khoekhoe, the same roots associated with different nominal designants yield the relevant meanings, and Quileute *pòxʷó·kʷoł* may be conceived of as variation of the association with musical instruments reported above, as it also means 'horn of a ship.' Similarly, the relevant Aguaruna and Huambisa terms (both apparently borrowed from Spanish) may also refer to a 'signal horn,' Bislama colexifies 'loudspeaker,' and Lesser Antillean Creole French colexifies 'motor horn,' alongside 'corn' and 'horny area.' Finally, in six languages, Noni, Swahili, Baruya, Toaripi, Lenakel and Samoan, the word for 'horn' may also refer to a 'tusk' (as of a pig or an elephant, for instance), and Hani

and Mandarin colexify 'horn' with 'corner,' among other meanings (see Matisoff 2004: 352 for further evidence from both within and outside Southeast Asia).

As already mentioned, there are many different associations by colexification, and a lot of them are found in only one sampled language. Hausa *k'aho* (presumably by provenience contiguity) also denotes a 'cigarette-holder,' but also "a kind of tattooing" and "a person who is either destitute or gives nothing away." Ngambay *gàjì*, alongside 'branch' and 'horn,' also means 'rubbish' inter alia. Baruya colexifies 'horn' with "curved moon, crescent moon" as well as "curve below the navel where skirts fit." Gurindji is the only language in the sample that uses the same term for 'horn' and 'root,' whereas Muna uses the term *tandu* not only to denote 'horn,' but also verbally with the meaning 'to butt, gore' inter alia. Abzakh Adyghe *bż(e)* can inter alia also refer to a drinking vessel such as a glass or cup, probably originally motivated by the former function of horns as drinking vessels, Badaga *kombu* also may refer to a "horn shape made from hill pavetta which is put inside the rooves [sic!] of houses after the uppaṭṭuva," and Greek *kéras* is also used to refer to a 'wing of an army.' Biloxi *ahi´ ~ ahe´ ~ ahĕ´ ~ he* is unusual in having the semantic range of "skin, fingernails, horn, hooves, scales of fish, bark of tree," and Upper Chehalis *wináẇ*, by perceptual similarity, also denotes a 'wedge.' Kiowa *gụądei* also means 'afterbirth.' The Santiago Mexquititlan Otomí term *ndäni* also denotes a 'ram,' whereas Pawnee's *paariiku*$^{?}$ clearly showcases meaning extension by metaphor: it also means 'colon' and 'banana.' Arabela *tuhuaja* is also used to refer to 'adornment' as well as 'crown.' *Kácho*, the Embera term for 'horn' can also mean 'slice' or 'piece' (it is in fact borrowed from Latin American Spanish *cacho* with the same semantic potential). Guaraní *tatî* also means 'prow,' while Ancash Quechua *waqra* can also refer to a 'useless thing.' Yanomámi *yõra* not only means 'horn,' but also 'fin.' The motivational history for Fijian *i leu* appears to be complex: *leu* is "to extract, as a thorn from flesh" and the derived nominal *i leu* in the first place means "the instrument for so doing: tweezers," and then also 'a small stick or needle' and finally, presumably by yet another metaphorical extension, also 'the horns of an animal.' A similar line of meaning extension may be hypothesized for *seru-na*, which is analyzable as 'to.comb-POSS.' Hawaiian *kiwi* also denotes the "horn" of the Kala fish (known in English as the Bluespine Unicornfish, *Naso Unicornis*), and furthermore any curved or bent object, and indeed, in an adjectival sense, also 'curved' and 'bent.' Rotuman *'ipesi* also denotes a particular type of 'flat wedge-shaped wooden spatula,' Samoan *seu* also means 'to steer,' 'stir, mix,' and 'to intercept,' White Hmong *kub* also means 'gold' (in this sense borrowed from Chinese) and 'to burn,' and, finally, Yay *kaw*^{1} also is used to refer to "a vine or creeper" and a 'protrusion' generally.

33. The Lagoon

Representation: 38%
Motivated: 51.5%
Thereof Analyzable: 23.5% Thereof Colexifying: 28.9%
Thereof by Contiguity: 16.7% Thereof by Similarity: 30.4%
Recurrent associated meanings: lake/pond, water/liquid, sea, puddle, swamp, tide, reef, round, big

Commonly, languages do not differentiate lexically between 'lake, pool' and 'lagoon,' using a single term for both referents. This is the case in Kwoma (colexifying also 'waterway, canal'), Tasmanian (Middle-Eastern and Southeastern), Toaripi, Yaqui (by a semianalyzable term containing an element meaning 'water'), Copainalá Zoque, Arabela, Aymara, Bora, Chayahuita, Cubeo (by the analyzable term *maca-jitabu* 'faeces-puddle;' compare colexification of 'lagoon' and 'puddle' in Copainalá Zoque and Ancash Quechua), Embera (where the meanings are associated with different genders), Guaraní, Huambisa, Miskito, Ancash and Imbabura Quechua, Tsafiki, Fijian, and Takia, while Greek has *limnothálassa* /límn-o-thálassa/ 'lake-STEM.FORMATIVE-sea,' Carrier *ʼa-pen-ḳet* 'fog-lake-on,' Hawaiian *loko kai* 'lake/inside sea,' and Mandarin *xie4-hu2* 'pour.down/fall-lake' and *jiao1-hu2* 'reef-lake.' Hawaiian has, like Mandarin, an association with 'reef' (*kai kohola* 'sea reef.flats,' which indeed can also refer to the 'shallow sea within a reef'). Greek, Mandarin, and Hawaiian are not the only languages associating 'lagoon' with 'sea': Nivkh has *kerq ŋalu* 'sea bay,' there is a semianalyzable term in Welsh, and San Mateo del Mar Huave and Kapingamarangi colexify the meanings directly (along with 'tide, salt water, salt' and 'to close up, shut' in the fomer language; for the association with 'tide,' compare Samoan *tai-tafola* 'tide-spread.out' and that Rotuman *maka* is glossed as "tidal flat, wide stretch of beach covered only at high tide" and also means 'to sing, dance'). Likewise, Yaqui is not alone with its association with 'water' or 'liquid': Yanomámi has *mono u* 'round.bodied liquid,' Takia parallely *you i-lanti* 'water 3SG-be.round,' Bororo *kuruga*, perhaps analyzable as *kuru-ga* 'liquid-go,' and Tsafiki *hua pipilú*, containing *hua* 'big' and *pi* 'water.' Other complex terms involving a constituent meaning 'water' are San Mateo del Mar Huave *waj-yow* 'neck/mane-water,' denoting a specific lagoon, and Tetun *bee-lihun* 'water-dam/tank.' Moreover, Guaraní directly colexifies 'water' (and 'river') and 'lagoon,' there is a semianalyzable terms with a constituent meaning 'water' in Guaraní, and a further semianalyzable term with a constituent meaning 'accumulation.' Kwoma and Yanomámi colexify 'lagoon' with 'swamp' (Yanomámi *wawëwawë* is a reduplicated form of *wawë* 'wide, empty'); Tetun *kolan* is also glossed as 'saltwater swamp' in parentheses. Bora colexifies 'lagoon' with 'parched arm of river,' and, conversely, Yir Yoront with "stretch of river where water remains in the dry season" as well as 'billabong.'

Other associations include: in Buin, the relevant term also colexifies the meaning "rainwater hole containing brackish water," Yoruba *ọ̀sà* also has a temporal meaning 'space of time, season, interval,'[5] and Kwoma *naba* can also refer to a swamp or a canal. Japanese *kata* also means 'shoulder' but has different prosodic structure from *kata* 'lagoon.' Bororo features two terms with unclear motivations for the concept: *baru-bo* is analyzable as 'sky-division' and *kuru-ga*, as mentioned above, perhaps as 'liquid-go.' Finally, Fijian *totobu*, meaning also 'deep place in center of stream,' is presumably reduplicated from *tobu*, meaning "pool in a river, bathing hole, well" inter alia, and Hawaiian *kua-ʻau* 'basin inside the reef, lagoon' is tentatively analyzable as 'back/windward-project.'

[5] The consulted source also has *adágún ibití omi òkun nṣàn sí* which is not a conventionalized Yoruba lexical item according to Joseph Atoyebi (p.c.).

34. The Lake

Representation: 82%
Motivated: 45%
Thereof Analyzable: 14.6% Thereof Colexifying: 31.7%
Thereof by Contiguity: 17.4% Thereof by Similarity: 10.6%
Recurrent associated meanings: puddle, lagoon, water, river, swamp, sea, spring/well, dam, big, round, rain, bay, inland, stand/stagnant

As discussed in § 6.2.2.5., bodies of water are sometimes not lexically distinguished from the substance 'water,' and this situation is also found for 'lake' (or 'pond') in five of the sampled languages, Berik, Itzaj, Bororo (by a register-specific term), Jarawara and Bislama (and is attested diachronically in Indo-European, evidenced by cognates meaning 'lake' in Avestan but 'water' in Sanskrit, Buck 1949: 38). In the case of Buli, Hausa, Ngambay, Rendille, Berik, Badaga, Khalkha, Bororo (by a term largely restricted to ritual language), Guaraní, Maxakalí, Jarawara, Piro, and Bislama 'river' is colexified ('perennial river' more specifically in Rendille), while there are, due to colexification of 'water' and 'river,' terms betraying this association by analyzable terms in Efik, Sko, Tsafiki, and Kaingang. Moreover, there is a semianalyzable term where the identifiable constituent can refer to both 'water' and 'river' in Kosarek Yale. In Itzaj and Jarawara, 'rain' is colexified (and similarly, 'rain water' is in Comanche). The Jarawara term *faha*, like Bislama *wota* and Itzaj *ja'*, is very general, the reference of which may include 'water' as a substance, 'river,' as well as 'lake' (and 'juice' and other meanings in Bislama).

More frequent are, however, associations with 'water' by way of morphologically complex terms. One pattern found in three languages of the Americas, Blackfoot, San Mateo del Mar Huave, and Tsafiki, is that the second element is 'big,' e.g. Blackfoot *ómahksíkimi* /omahksi-ikimi/ 'older/large-liquid' (note, however, that the San Mateo del Mar Huave term *nadam yow ngo mawaag* contains the unknown element *mawaag* plus, apparently, the negating morpheme *ngo*, and compare also the redundant complex term *tɔ̀ku ba* 'big water/lake' in Ngambay, and the same situation in Itzaj). In three sampled languages, Yanomámi, Rotuman, and Takia, the additional element is 'round,' e.g. Yanomámi *mono u* 'round.bodied liquid' (compare in addition Lengua *yakyengyiam* 'lake' and *yakyeyi* 'be round). Other complex terms for 'lake' where one of the constituents is 'water' include Efik *mkpö-diök'hö-möñ* 'thing-be.placed water' and *(ëbiët) udiökhömöñ* /ëbiët u-diök'hö-'möñ/ '(place) NMLZ-stand.level water,' Sko *pato* /pa-tó/ 'water-inside,' Abzakh Adyghe *psə-wɔć°ə-ɣe* 'water-stand-PAST,' Cheyenne *tsé-sééha mâhpe* 'that.which.is-spread.out water,' Tetun *bee-lihun* 'water-dam/tank' ('lake' and 'dam' are colexified in Badaga and Chickasaw) and Rotuman *tạn hăe* /tạnu hăe/ 'water contain' (this term also means 'puddle' and 'bay, inlet,' which latter association is shared with Nez Perce). Cahuilla *pàl múyeqalet* contains *pál* 'water, river' and *-múye-* 'fill up;' the literal translation provided by the lexicographer is "water which fills up." In addition, Maxakalí *puxhep* contains *pux* 'pour out' and *hep* 'blood, sap, liquid,' and *kõnãgkox* is analyzable as /kõnã'ãg-kox/ 'water-hole.' Furthermore, there is a wealth of semianalyzable terms for 'lake' where the element 'water' is discerni-

ble, but the full morphological structure cannot be eludicated. This is the case in Noni, Kosarek Yale, Kiowa, Comanche, Yaqui, Yuki, and Samoan.

Associations with the 'sea' are also found, either by colexification (seven languages, Ngambay, Khalkha, Nez Perce, Tuscarora, Tehuelche, Yanomámi, and Hawaiian; diachronic shift is well-attested in Germanic, Buck 1949: 38) or by morphologically complex terms. This is found in the Japanese term for a 'big lake' (*mizu-umi* 'water-sea'), but is also found in two languages of Africa which have interestingly a very similar conceptualization strategy. In Bakueri, the 'lake' is called *mmána mmánja* 'child sea' or *ngúmá mmánja* 'younger sea,' and in Khoekhoe *hurirob*, a term used in Bible translations, consists of *huri* 'sea' followed by the diminutive suffix *-ro* and the nominal designant *–b*. Moreover, there is a semianalyzable term in Hani where the identifiable constituent means 'sea, ocean' as well as 'to soak in water.' In connection with the metaphorical transfer of a 'child' as a prototypically small referent (Jurafsky 1996), the Waris term *polomb* deserves some discussion. It obviously contains *pol* 'liquid,' and the second element looks like a truncated form of *ombol* 'son.' There is a parallel for this from the same area: Baruya features the term *budaaya* for a 'small lake' which is formally redundant, since the constituent *bunya* alone can already be used to refer to a lake, but the other constituent of this compound is *taaya'* 'girl.' A further analyzable term is Yanomámi *wawëwawë*, the reduplication base of which appears to be *wawë* 'wide, empty'

Exclusively by colexification, associations between 'lake' or 'pond' and 'swamp' are found in eleven of the sampled languages, namely Buli (where the interpretation of the relevant term as 'lake, pond' is rare), Ngambay, Gurindji (where the term also denotes a specific lake and "something in ceremony"), Kwoma, Basque, Ket, Chickasaw, Lesser Antillean Creole French, Kaingang (by the analyzable term *óré ki goj* 'mud in water/river'), Pawnee, Yanomámi (by the term *wawëwawë*, reduplicated from *wawë* 'wide, empty), and Hani (and see Buck 1949: 38 for Indo-European evidence), with 'lagoon' in 21 languages, Kwoma, Tasmanian (Middle-Eastern and Southeastern), Toaripi, Yaqui, Copainalá Zoque, Arabela, Aymara, Bora, Chayahuita, Cubeo, Embera (where the meanings are associated with different genders), Guaraní, Huambisa, Miskito, Piro, Ancash and Imbabura Quechua, Tsafiki, Yanomámi, Fijian, and Takia, and with 'spring' or 'well' in Buli, Wintu, Yuki, Cayapa, Kaingang, and Maxakalí (by the analyzable term *kõnãgkox* /kõnã'ãg-kox/ 'water-hole'). Given the possible variation in the size of lakes, from very small to very large, and the fluid boundaries between a lake and a pool down to a small pond, the meaning 'puddle' is also colexified in 24 of the sampled languages, namely Efik, Baruya, Buin, Kyaka, Rotokas, Kosarek Yale, Abzakh Adyghe, Cheyenne, Haida, Lesser Antillean Creole French (also colexifiying 'mast, pole'), Tuscarora, Copainalá Zoque, Bora, Cashinahua, Guaraní, Ancash Quechua, Hawaiian, Lenkakel, Rotuman (by the analyzable terms mentioned above), Sedang, and Yay. Cubeo *macajitabʉ* is, however, peculiar, apparently consisting of *maca* 'faeces' and *jitabʉ* 'puddle.'

In two Polynesian languages spoken on small islands, Hawaiian and Fijian, the meaning 'inland' or 'interior' is also colexified; in Hawaiian, there is an optional complex term *loko wai* 'interior/lake water.'

Other associations are also found cross-linguistically. Buli *biung* is primarily used to refer to a "watering hole that dries up quickly" and rarely assumes the meaning 'lake, pond.' Hausa *tabki ~ tafki* also refers particularly to "water in a borrow-pit" (it is also an "exclamation of astonishment at bigness") and *bingi* also to a " 'rough-coated' fowl" as well as "[a]ny large donkey" inter alia. Swahili *ziwa* also means 'breast,' Kwoma *naba* is also used to refer to a 'waterway' and 'canal,' Kyaka colexifies 'lake' with 'shallow water,' and Ngaanyatjarra *murrkungu*, denoting a 'salt lake' specifically, is by metonymy also used for 'salty soil,' while *parntu* can also refer to 'salt' itself. Badaga *aḷḷa ~ haḷḷa* also may refer to a "bottomland, lowest spot, depression" and 'floodwater.' Khalkha *naɣur* can also refer to a "dry lake bed." Welsh *llyn* is also used with the meaning 'liqour, drink' and *llwch*, an obsolete term for 'lake,' also means 'dust, powder.' Kolyma Yukaghir *jalɣil* is also the name for the tambourine of a Shaman, and *ńorol'* colexifies 'moss.' Kashaya colexifies 'slick,' Lakhota 'I went,' the reduplication base of Lake Miwok *pólpol* means 'float, flood' (although *pólpol* may be a loanword from Cache Creek Patwin as a whole), and Wintu *λul*, also denoting a 'water hole,' may be etymologically connected to a word for 'to bubble' (there is also the term *sa·wal* for a mountain spring with mythological significance). Cavineña *bei* is also the name of the 'anteater,' Hawaiian *loko* also means 'in, inside' and may refer to the 'internal organs' inter alia, while *moana* is, alongside its usage to denote both 'lake' and 'ocean,' also applied to name a "campground, consultation place for chiefs" (the common denominator apparently being that both are an expanse). Bwe Karen *nò* also denotes a 'moat' and a "written musical note" (perhaps, as the source suggests, due to English influence), while Lenakel *nɨsíu* also denotes a particular lake, namely Lake Siwi on Tanna Island, Vanuatu. Mandarin *hu2* also means "bottle-gourd, flask, teapot" (in both meanings going back to Early Middle Chinese ɣɔ, Pulleyblank 1991: 126), Sedang colexifies 'naked,' and by another term 'ripe, red,' and Lesser Antillean Creole French *lak* also may refer figuratively to a "slackening of control.'

35. The Lightning

Representation: 93 %
Motivated: 39.9%
Therof Analyzable: 22.0% Thereof Colexifying: 18.4%
Thereof by Contiguity: 12.6% Thereof by Similarity: 22.0%
Recurrent associated meanings: thunder, gleam/lighten/shine, electricity, light, spark, spirit/god, rain, fire, thunderstorm, sky, eye, telegraph, flame, arrow, blink, sword

Metaphorical denominations for 'lightning' (or 'thunderbolt') are common in the world's languages, and some of them have a remarkable areal distribution (see § 6.4.3.). In Eurasian languages (represented by Khalkha and Kildin Saami in the sample), complex terms for 'lightning' using the source concepts 'thunder' and 'arrow' are found (*ajungɣa jin sumu* 'thunder GEN arrow' and *tīr'm'es'-ŋull* 'thunder-arrow' respectively). Another pattern predominantly found in Eastern Eurasia (Ket and Kolyma Yukaghir), but also in languages of the American Northwest (Central Yup'ik and Kashaya) is a lexico-semantic association

with 'fire:' Ket has *ekkinna boʔk*, analyzable as /ekkin-na boʔk/ 'thunder-3POSS.AN.PL fire,' Kolyma Yukaghir *jedun-ločil* 'thunder-fire,' Kashaya *maʔƙala ʔoho* 'thunder fire,' and Central Yup'ik *kenerpallak* /keneq-pallag/ 'fire-intensively;' there are semianalyzable terms where 'fire' is the meaning of the identifiable constituent in Chukchi and Kildin Saami (the association is also weakly attested in Indo-European, Buck 1949: 57). Areality with respect to the meaning 'lightning' is also found in languages of Southeast Asia: the source concept 'blink' is encountered in Manange and White Hmong (Manange *tipli-1k^{h}ʌ* 'blink-come' and White Hmong *xob laim* 'xob blink,' see Bauer 1992 for further discussion). White Hmong, by way of having a term referring to a spirit named *Xob*, also participates in another putative Southeast Asian pattern, which is precisely characterized by making reference to a spirit: Yay has *pya^{3} θaw^{5}* 'pya^{3} throw' and *pya^{3} ta^{4} yaaŋ^{6}* 'pya^{3} pull/draw sword,' *pya^{3}* being the spirit causing thunder and lightning (note in this context also Japanese *kami-nar-i* 'god-sound-NR,' as well as the fact that Burarra *andarrbaykarda* is also the name of a lightning spirit, and compare also the association between 'lightning' and 'Thor's hammer' in Old Norse and with 'Perkun's hammer' in Lettic). In turn, the association with 'sword' in Yay is also likely to be part of an areal pattern of Southeast Asia, the evidence being that Sedang has *cháng tơro* 'sword/men's tool thunder.' One of the source concepts in two languages of New Guinea, Meyah and One, is 'cloud;' the Meyah term is *mocgój efésa* 'fog/cloud flash,' the One term is semianalyzable only. Two languages, Tsafiki and Tetun, have an association between 'lightning' and 'flame' (*cunta pinda* 'thunder lightning/flame' and *rai-lakan* 'land flame' respectively; note also Yaqui *yuku be'oktia* /yuku be'ok-t(e)-ia/ 'rain-lightning-INTR-NMLZ' which also means "to take out the tongue several times" and the semantic connection between 'tongue' and 'flame' reported in section 22; Yaqui also has the term *yuku jimaa-ri* 'rain throw-RES').

Unsurprisingly, 'lightning' is also associated with meanings such as 'to gleam,' 'to lighten,' 'to shine,' or the like (common also in Indo-European, Buck 1949: 56-57). Hausa *walk'iya* also means 'glossiness' alongside 'lightning,' Toaripi *kevaro* also means, often reduplicated, "the shine, gleam on the leaves of plants and trees, or on a person's skin; the flash or sparkle of anything bright." Associations like these are also found by colexification in Abzakh Adyghe, Arabela, Bororo, and Lenakel, while Efik and Toba have derived terms, Sora *'kila:igum* is analyzable as /'kila:i-gum/ 'shine.brilliantly/dazzle-rain' (there is another semianalyzable term with a constituent meaning 'rain' in Sora), Carib *kapekape* is reduplicated from *kape* "smoothness, gleam," Guaraní *aravera* is analyzable as /ára-vera/ 'sky-brilliance' (Guaraní also has *aratiri* /ára-tiri/ 'sky-crack;' for the association with 'sky,' compare also Kwoma *neer hopo* 'sky snake' and Hani *aoq-miaovq miaovq* sky-burn RED), and Tetun *rai-nabilan* as 'land-shine.' There is a semianalyzable term where the identifiable constituent means 'to shine' in Hani. Another association that is akin to that just discussed is that with 'light,' occurring in six sampled languages. In Chukchi, *jənqeryen* (containing *jən* 'fire') means both 'light of fire' and 'lightning,' and in Rendille 'lightning' is colexified with 'flash(es) of light' generally. Lenakel *nasiapumelaan* is derived from *asiapumel* 'to lighten, flash,' which in turn contains *asia* "[m]ake a light, make a torch (from coconut fronds)." If the association is by analyzability of the lexical type, there is variation in the semantics of the second element: in Berik and San Lucas Quiaviní Zapotec, it is

'thunder' (*iris naf* 'thunder light' and *x:cha' x:tèe' bzìu'* 'light GEN thunder' respectively); in Maxakalí, it is 'rain' (*tex yãnãm* 'rain light'). In fact, as already seen in the Yaqui example cited above, 'rain' is also a meaning that is frequently associated with 'lightning.' Alongside Sora, Yaqui, and Maxakalí, which were already discussed, the pattern is also found in Mbum, where 'lightning' is either *sàk à mbàm* 'tear GEN rain' or *máà-mbàm* 'mother-rain.' An association with the 'eye' is found in three languages of North America, Ineseño Chumash, Nez Perce, and Tuscarora, although it is doubtful whether they represent the same conceptualization strategy. In Ineseño Chumash, the 'lightning' is called *štix a soxk'on* "the eye of the thundercloud." In contrast, in Nez Perce, it is called *taqasaʔyóˑxoʔt* /teqe-se-ʔiyóˑxoʔ-t/ 'suddenly-eye-watch,' suggesting that this term is not metaphorical in nature, but rather refers to the sudden perception of light by the human eye. The same is true of Tuscarora *newatkahréhnariˑks*, containing the roots *-kah(r)-* 'eye' and *-rik-* 'bite.' In Lakhota, a literal translation of the word for 'lightning,' *wakíyątųwąpi*, would be 'the thunderbirds are looking' (compare the notion of the thunderbird in North America mentioned in § 6.4.3.15.5.). Further complex terms for which the internal structure is not entirely clear include Kiowa's *bǫųębʜheipàgyʜ* which appears to contain the root *bǫųę-* 'transparent,' Sko's *hénghèng* (putative reduplication base: *hèng* 'fart') and Bororo's *baigabe* (compare *baiga* 'Bororo bow' and *be* 'excrement').

An interesting conceptualization of the meaning 'lightning' is found in Haida, involving verbal classifiers and the word for the colour 'red:' *sri q'asda* contains *sri* 'to be red' and the classifier *q'a* for loud sounds (or possibly for large twodimensional surfaces) and *srid raa7uhlda* the classifier *raa* for flashing light and *7uhlda* 'to blink eyes' (for which compare the association with 'eye' in other North American languages discossed above; note also that Comanche *ekakwitseʔe* may contain *eka-* 'red').

As already seen in various examples, complex terms for 'lightning' unsurprisingly frequently are made up in part of terms for 'thunder.' Other complex terms with that structure other than those already mentioned are Japanese *raku-rai* 'fall-thunder,' Carrier *tı̨̂tni ełkreš* 'thunder emit.flashes,' and Cayapa *cuidya pi'queno* 'thunder little' (obviously of relatively recent vintage due to the presence of the loan from Spanish). However, 'thunder' and 'lightning' are also frequently associated by colexification, which is the case in 21 sampled languages: Buli, Ngambay, Rendille, Yoruba, Kwoma, Tasmanian (dubiously), Abzakh Adyghe, Japanese, Itzaj, Xicotepec de Juárez Totonac, Yana, Yaqui, Abipón, Arabela, Aymara, Hupda, Jarawara, Tehuelche, Bislama, and Bwe Karen. In Itzaj, the meaning 'lightning' can be singled out by the complex term *jatz' chaak* 'whip thunder/lightning,' and in Takia by the verbal terms *weil i-fni* 'thunder/lightning 3SG-hit' and *weil i-raklawi da* 'thunder/lightning 3SG-wink IPFV' (for this association compare Efik *ekepkep*, derived from *kep* meaning 'to corruscate, flash, lighten,' but also 'to wink'). There is a semianalyzable term with an identifiable constituent with the meaning 'thunder' in San Lucas Quiaviní Zapotec, and in three further languages, Ngambay, Burarra and Tetun, 'thunderstorm' or 'electrical storm' is additionally colexified.

By colexification, associations with 'spark' are found in six sampled languages, namely Abzakh Adyghe, Badaga, Abipón, Bora, Ancash Quechua and Lesser Antillean Creole French. Yanomámi colexifies 'lightning' with 'electric spark' particularly, and indeed, a

relatively common pattern of semantic extension is to use the term for 'lightning' also for 'electricity.' This is found in Swahili, Basque, Khalkha, Nez Perce, Pawnee, Hupda, Hawaiian, and Samoan (the relevant Khalkha and Nez Perce terms colexify also 'telegraph').

Other associations include: the general meaning of Buli *ngmoruk* is 'rain,' and it rarely refers to 'thunder' and 'lightning.' Ngambay *ndàngè* also means 'to scold someone' and 'hurry,' and *tèel* also means 'suffering, pain' and 'announcement.' Muna has *bhibhito*, containing *bhibhi* 'to quiver, shake,' Nunggubuyu *-marawadja-*, containing *-$w_2$adja-* 'to hit, kill,' and for Rotokas *parakau,* compare *paraka* 'wide, spread out,' 'width' and the classifier for narrow objects *ua*. Colexification with 'hailstone' is found in Basque, and Japanese has *inazuma*, analyzable as /ine-tsuma/ 'rice-spouse.' In Khalkha, *vacir ~ vcir ~ ocir* is also the name of a sacred instrument in Buddhist ceremonies, Cheyenne *vovó'ho'kâsé'há* contains *vovó'h* 'spotted white' and also means 'to flash repeatedly,' Itzaj *lemlem ~ lenlem* appears to be reduplicated from *lem* 'calm,' while in Jarawara, the same term is uniquely (with respect to the sample) also used for the 'sun' (and by extension also 'clock,' see discussion in section 79) and both 'lightning' and 'thunder.' Wichí has *pelhach'e* /pelhay-ch'e/ 'storm-stick,' Bislama *laetning* also means 'in excess' and 'very quickly,' and Fijian *liva* also 'lift, lever' (in this sense due to borrowing from English). Kapingamarangi *ila* is also the term for a 'mole on the skin' or a 'maggot,' Rotuman *mere* also means 'to criticize, find fault with,' and colexification with 'wheel' in Samoan is almost certainly secondary due to English influence - the relevant term is *uila.*

Finally, it should be mentioned that in many languages, the semantics of 'lightning' is primarily or exclusively encoded verbally and the noun being derived from it, due to the temporal instability of the concept and its essentially event-like nature. This is for instance the case in Khoekhoe (*napa-b ~ tapa-b ~ lapa-b* 'to.strike.as.lightning-3SG.MASC) and in Laz, where the verb *divalai* is used to refer to the event of lightning.

36. The Meteoroid/Shooting Star

Representation: 48%
Motivated: 61.5%
Thereof Analyzable: 55.9% Thereof Colexifying: 5.6%
Thereof by Contiguity: 5.3% Thereof by Similarity: 54.3%
Recurrent associated meanings: star, tail, faeces, fall, fly, fire, run, spark

Terms for the meteoroid (for which 'comet' was accepted as a proxy if no other terms could be retrieved from the sources) are most frequently of the lexical type, one of the consituents being 'star' (or very rarely, a term with a fire-related meanings such as 'spark,' see below for further possibilities) and the other standing in a metaphorical relation with the target concept 'meteoroid.' In some languages, the second constituent is verbal and means either 'to fall,' 'to fly,' or less commonly 'to run.' For instance, One has *leila fanta yolu* 'star ?? fall,' a pattern also found in Khoekhoe, Toaripi, Cheyenne, Yuki, Bislama, and Tetun, Yuki has *mančipą·se ṭiʔ-ik* 'star fly-??,' a pattern also found in Khalkha, Hawaiian and Samoan, and Chickasaw *fochik malili-'* 'star run-NMLZ,' a pattern also found in Abzakh Adyghe and Kapingamarangi. Even clearer metaphorical denominations are found when the

second constituent is nominal in nature: recurrent patterns of this kind include a metaphorical comparison with 'tail,' as in Yoruba *ìràwọ̀onírù* /ìràwọ̀-oní-ìrù/ 'star-owner-tail' and also in Ineseño Chumash and Rama (where an additional element meaning 'sending' is present'), or, a little more commonly, 'faeces,' as in Highland Chontal *łifay galxamna* 'faeces-star,' and also in Haida, Central Yup'ik (where 'meteor' is colexified with 'puffball,' since "meteors are traditionally said to turn into puffballs when they land"), Toba (where the relevant term colexifies 'meteoroid' with 'mushroom,' Toba also has *to'olloxoic,* apparently derived from *to'olloxoi* 'mushroom' by means of the masculine suffix -c; for the association between 'mushroom' and 'faeces,' see section 41 and § 6.2.3.3.), Sedang, and Tetun (where an additional element meaning 'hurl' is present). Further complex terms involving 'star' are Basque *izar koloka* 'star loose,' Khalkha *suɣunaɣ odu(n)* 'column.of.smoke star,' Carrier *sem-thełṣek* 'star-uses.to.go.off,' Upper Chehalis *scakʷútwaln tat łačis*, containing *łačis* 'star' and the reciprocal marker *-twal*, presumably among other morphemes, Kashaya *qʰa·moż šuṭʰuhṭadu,* analyzable as /qʰa·moż šu-hṭʰuṭ-ciᴼd-w/ 'star by.pulling-pieces.come.off.bigger.object-DUR-ABS,' Bororo *ikuieje ukigareu*, containing *ikuieje* 'star' and *kigareu* 'adorned,' Miskito *slilma dakwi ba*, consisting of *slilma* 'star,' *dakwaia* 'break' and the demonstrative element *ba*, Piro *katagiri psojite* 'star fragment,' and Fijian *kalokalo cavu* 'star eradicate' (with *cavu* also having other meanings, among them 'adorned, highly decorated'). Moreover, Rotuman *hef sàl'ạk mala* contains *hefu* 'star' and *mala* meaning 'red hot' but also denoting a red kind of belt worn by high chiefs, and a further San Mateo del Mar Huave term for the 'meteoroid' is *nandaab ocas* 'burnt star.' Colexification with 'star' itself (as well as 'planet') is encountered in Toaripi and Sora, while in Khoekhoe, the same term suffixed with different nominal designants yields the meanings 'star' and 'comet' respectively; in both cases, the relevant terms are derived from a verb meaning 'to blink, twinkle.' In Bislama, *sta* is glossed as "any heavenly body (e.g. moon, star, meteorite)." There are semianalyzable terms where the identifiable constituent is 'star' in Mbum, Sko, Biloxi, and Chayahuita. None of the abovementioned patterns has a clear areal hotspot of occurrence, rather, each one recurs in many different areas of the world.

There are also variants of some of the associations in which, rather than 'star,' some other meaning figures. Guaraní has *jagua-veve* 'dog-fly,' and Tuscarora *nekačis-nahkwá?nę?* contains the roots *-čiʻsn-* 'spark' and *-ne'nę-* 'fly' (for which compare Hani *miqzaq miqseil ~ aqzaq miqseil* 'fire spark;' *miqseil* is also "a woman who is dissatisfied with her marriage"). Variants of the denomination via 'tail' are San Mateo del Mar Huave *miwiül ix* 'tail rock.iguana' and Dongolese Nubian *káǧ-n-ɛ̄u* 'horse-GEN-tail' (indeed, both terms can also be interpreted literally and refer to an iguana and horse tail respectively). Badaga is the only language in the sample where this association is realized by colexification (also colexifying "slender means" and "anything meagre"). Lake Miwok *wikíiwiki* is a reduplicated version of *wíki* 'fire,' while Itzaj colexifies 'shooting star' with 'fire' and other meanings directly. Guaraní *yvarata* is analyzable as /yva-tata/ 'fruit-fire' (there is also a semianalyzable term where the identifiable constituent is 'sky').

Another strategy is found in Middle-Eastern Tasmanian, where *pökarīt'ę* appears to contain a constituent meaning 'ghost' (compare the annotation for Kwoma *maway*: "Shooting stars are thought to be the souls of outstanding warriors who have died. At the

second, or final, burial of such a man his soul, in anthropomorphic form, flies through the air holding a burning coconut frond torch in its hand, which is the light seen in the sky, towards one of the many large lagoons that surround the Washkuk hills"). Great Andamanese *chàugalalachōinga* is somwehat similar in that it makes reference to the 'spirit' (*chàugala*); *chōinga* means 'light of torch.'

Other unique denominations are found in Efik, where *inöntanta'fioñ* contains *i'nö* 'thief' and *ika'* is derived from *ka* 'to go.' Sahu has *detere'e ma ngi'di* 'tooth POSS thunder,' and Welsh *maen mellt* 'stone lightning' alongside *awyrfaen* /awyr-maen/ 'air/sky-stone.' Basque *meteor* is also glossed as 'atmospheric phenomenon.' A literal translation of Kiliwa *xsmii??aawpmaay* is 'Xsmii's fiery urine' (Xsmii is the name of a constellation; compare the associations with 'faeces' discussed above). Wintu *nomλeyna·s* is literally 'going west.' For Abipón *neiak*, compare *eiagaiag- ~ -eiagaiak-* 'walk, travel.' Miskito has *imyula mabra* 'lightning egg,' while Piro *gijrukachri* appears to contain *gijru* 'shaft, handle.' Yanomámi *thoru wakë* consists of *thoru*, the name of a plant species, and *wakë* 'red,' and is at the same time the name of the fire spirit. Patterns of colexification include that with 'spear' and "standing upright on the hands" in Hausa, "a flaming arrow, as a signal, or for setting fire to enemy villages" in Buin, that with 'angel' in Rotokas (the term, *purapurapato*, is likely derived from the verb *pura-* 'to contract something, say, make, do, create' and the derivational element *-pa*), that with 'Johnny Jump ups' (a species of violets) and 'flowers' generally in Wintu, that with 'stick, crop' in Aguaruna, and that with 'thunder without clouds' in Arabela.

37. The Milky Way

Representation: 38%
Motivated: 70.3%
Thereof Analyzable 65.3% Thereof Colexifying: 5.9%
Thereof by Contiguity: 1.8% Thereof by Similarity: 66.5%
Recurrent associated meanings: trail/road/street, star, sky, river, eel, ghost, gird, backbone, tapir

For terms for the Milky Way (or 'galaxy,' which was accepted as a proxy), it is the rule rather than the exception to be motivated. Apparent exceptions are found in Katcha, Khoekhoe, San Lucas Quiaviní Zapotec, Nunggubuyu, Great Andamanese and Kapingamarangi, for which no analyzability can be safely diagnosed on the basis of the source and no further meanings are stated. Lexico-semantic associations are often culture-specific and bear mythological connections. For instance, Rendille *íntí waraába abártiís kájiité* is literally translated as 'the place where the hyena dragged his mother' and comes from a Rendille children's story of the same name, and in Ket the name for the Milky Way is *Albakaŋ*, analyzable as /alba-kàŋ/ 'Alba-way,' Alba being a Ket mythological hero (Andrej Nefedov p.c.). In spite of the highly language-specific associations for the Milky Way, there are nevertheless some general naming tendencies, and the Ket example already provides an example of this: in fact, a recurring denominational pattern for the Milky Way found also in many well-known European languages is that involving various travel paths,

such as 'trail,' 'road,' 'street,' and also in a few cases 'river,' with the second element usually language-specific. Thus, Muna has *sala waghua*, containing *sala* 'path, trail' and presumably *ghua* 'gray hair,' Blackfoot has *makóyoohsokoyi* /makóyi-mohsokó-yi/ 'wolf-road-INAN.SG,' Upper Chehalis *sxai-yoxoɬ* 'trail-poor.people,' Cheyenne *Áméó'o* is analyzable as /ame-meo'o/ 'pemmican-way' and *séotsé-méó'o* perhaps as 'corpse-road' (the source notes difficulties with the precise translation), Chickasaw *Ofi' Tohbi' Ihina' ~ Ofi' Tohbihina'* as /ofi' tohbi-' im-hina'/ 'dog be.white/be.pale-NMLZ DAT-road,' Itzaj has *säk-b'ej* 'white/gray-trail/road' (colexifying 'highway'), Lakhota *wanáǧitȟačhą́ku* /wanáği-tha-čhą́kú/ 'ghost-??-road,' Nez Perce *ċewċéʷw-nim ʔískit* is 'ghost-POSS trail,' Wintu *qanal yemer* 'be.open.wide road' and *ḳaqal yemer* 'oblivion road.' Chayahuita *chimirin ira* is 'death way,' Kaingang *krĩg japry* 'star way' (there is a further semianalyzable term with *krĩg* that is also the name of a particular star), Miskito *swara bila* 'eel way' (*bila* also has other meanings alongside 'way'), Rama *núnik kás aríra*, a literal translation of which is 'cloud street' (although *aríra* is glossed as 'string, fold' on its own), Wayampi *tapiʔi-la-pɛ* is 'tapir-of-way' (a further variant being *tapiʔi-ɬipo* 'tapir-track,' compare Guaraní *mborevirape*, also containing *mborevi* 'tapir' and *tape* 'way'), Fijian *sala-ni-cagi* 'road-POSS-wind/air,' Sedang *tróang hơlóng* 'road star' (alongside the alternative term *hơlóng nhéng* 'star mirror').

As the examples show, the Milky Way is often associated also with otherwordly phenomena, such as 'ghosts' in Lakhota and Nez Perce and 'death' in Chayahuita, but by no means necessarily so. Nor is it the case that the association with travel paths is always by morphologically complex terms, Hausa is an example of a language with colexification (here, also 'channel,' 'intermediary,' and 'beehive' are colexified inter alia). A further case of colexification is perhaps Tasmanian (Plomley 1976: 408), and Muna has a semianalyzable term.

Yir Yoront and Ancash Quechua colexify 'Milky Way' with 'river' directly (in Ancash Quechua, there is the optional complex term *paqas mayu* 'night river'), Chukchi has *cəɣej-weem* 'sandy-river' and Tetun *mota-klakatak* 'river-reflection/image' and *mota-leten* 'river-top/summit.' In Khalkha, an association with a 'gird' is found by a morphologically complex term (*oγturγui jin byse* 'sky GEN girdle' alongside *tngri jin ojudal* 'sky GEN seam'), and one of the glosses of Hawaiian *kau* is 'to gird.' The association with 'milk' (the tertium comparationis obviously being the whiteness) is restricted to languages of Europe in the sample: Basque has *esne-bide* 'milk-way' (as well as the alternative term *santiago-bide* 'Santiago-way') and Greek a synchronically semianalyzable term. However, there are also languages outside Europe in which associations with whiteness seem to occur. For instance, Sahu has *ka'e ma geolo* /ka'e ma ge'olo/ 'drink/palm.wine POSS foam,' Kiliwa *ʔ+maʔi=ny-ʔ+phuuy* 'DN+sky=POSS-DN+smoke,' Bororo *kuiejedoge eerugudu*, containing the words for 'star' and 'ash' (there is also the alternative term *ikuiejedoge erugudu* containing *ikuieje* 'star' and *erugu* 'see'), San Mateo del Mar Huave *Minajndot Oleaj Micawüy Santiago*, containing *najndot* 'dust' and *cawüy* 'horse.' Note also Xicotepec de Juárez Totonac *i'xtej stā'ná' kaxtaj* (*kaxtaj*, 'lime') alongside the already mentioned Chickasaw and Itzaj terms.

In Hawaiian, the Milky Way is either *iʻa*, the basic meaning of which is 'fish' (it can also refer to an 'eel' inter alia specifically, compare the Miskito association with 'eel' mentioned above), or *iʻa-lele-i-aka* 'fish-jump-in-shadow' (and there is also the term *hōkū-*

nohoaupuni ‘star-rule’). As seen from the examples discussed so far, ‘star’ is an as frequent as obvious meaning of constituents of complex terms. Alongside the languages already mentioned, an element with this meaning also features in Sora (*ə-oːn-ˈtuˈi-ən* ‘POSS-child-star-N.SFX’), Nuuchahnulth (*ƛaaƛusaʔuɬ*, analyzable as /ƛaƛus-aʔuɬ/ ‘star-along’), and Hani (*aqgeel caqkov*, presumably analyzable as ‘star ditch.around.house,’ although ‘star’ is more fully *aqgeel alsiq* or *aqgeelsiq*).

In two languages of North America, Ineseño Chumash and Kiowa, an association with the ‘back bone’ is found: *šnokok’ a saxiyi’* “the spine of the night” and *tᵻ̨’-gǫ̨ṃtʻǫ̨* ‘star-backbone’ respectively (in Ineseño Chumash, an alternative term is *’alšuyopoš* “the piñon gatherer”).[6] Tuscarora *yurę́hyáhuks*, containing the roots *-rę́hy-* ‘sky’ and *-huk-* ‘light up,’ can also refer to a ‘rainbow’ and the ‘Aurora Borealis,’ and the Yuki term *miṭ ʔon ƙaw* also betrays an association with ‘sky’: it is analyzable as /miṭ ʔon ƙaw/ ‘sky land light.’ Likewise, Hawaiian *lā-lani* might be analyzable as ‘sun/day-sky’ (though both putative constituents also have many other meanings).

Other associations include: in Welsh, the Milky Way bears the name of mythological castles: *caer Wydion* and *caer Arianrhod* (*caer* ‘castle’), and in Carrier it is called *yą̲-ḳe-tsîlkrai* ‘sky-one-run.over.’ Pawnee *Rakiiraruhuuturuuhat* contains *uhak* ‘to pass in a line,’ and in Central Yup’ik *Tanglurallret* contains *tangluq* ‘snowshoe;’ it is so called because “because in legend it is the snowshoe trail of the raven.” Associations by colexification are, as noted above, much rarer, though not nonexistent: Hausa, by another term than that mentioned above, also colexifies “[t]he top of the head of a horse, between the ears” and “[t]he top of the occiput of human beings” inter alia. The Ngaanyatjarra terms *tjukal(pa)* and *yintirri* both also mean ‘tree ladder,’ Nunggubuyu *burumburunᵍa* also “galaxy, cluster of stars,” and Khalkha *mecid* also denotes ‘monkeys’ or ‘apes’ (formally it is the plural of *meci(n) ~ beci(n)* ‘ape, monkey, ninth year in twelve-year cycle’). Hawaiian *kau* has, alongside ‘Milky Way’ and ‘fish, eel,’ also the meaning ‘to place, put’ and is the “name of a star in the northern sky that served as guide to mariners” inter alia, and there is also the term *lele-aka* ‘fly-shadow’ and a further term containing *lele* which colexifies ‘shark-sucker’ and ‘remora’ inter alia. For Kapingamarangi *ganiwa* compare *gani* ‘penis’? The relevant Lenakel term contains a constituent meaning ‘girls.’

38. The Moon

Representation: 100%
Motivated: 67.1%
Thereof Analyzable: 6.5% Thereof Colexifying: 60.6%
Thereof by Contiguity: 52.5% Thereof by Similarity: 5.2%
Recurrent associated meanings: month, sun, moonlight, season, night, albino, menstrual period, light, snail, bright, white, torch

[6] In the English-Kiowa section, the term is given as *tą̲’-gǫ̨ṃtʻǫ̨* in which case it would be ‘sibling-backbone;’ this is likely a printing error, since under the lemma for *gǫ̨ṃtʻǫ̨* the term for ‘Milky Way’ is stated to be *tᵻ̨’-gǫ̨ṃtʻǫ̨*.

Many of the analyzable terms found cross-linguistically are due to the fact that many languages of the Americas, but also some in Siberia, lack true lexical differentiation for 'sun' and 'moon,' and have a single term for the two luminaries (found in 17 sampled languages all in all: Blackfoot, Carrier, Upper Chehalis, Cheyenne, Chickasaw, Lakhota, Nez Perce, Nuuchahnulth, Quileute, Tuscarora (according to an older source incorporated into the consulted source), Wappo, Wintu, Bora, Cayapa, Cubeo, Hupda, Macaguán, and possibly Tehuelche, where the relevant terms are very similar segmentally. In some of these languages, 'watch' and/or 'calendar' is also colexified due to the 'sun'-reading, see section 60 for fuller discussion, and sometimes there is the possibility of optional disambiguation. Another possibility is morphologically complex terms for the 'moon' containing the respective word for 'sun,' for instance Lake Miwok *káwul híi* 'night sun.' The latter situation is encountered also in two other sampled languages, Kolyma Yukaghir (*emin-pu:gu* 'night-sun') and Maxakalí (*mãyõn-hex* 'sun-female/man's.sister,' for which compare Rama *tukán kumá* 'moon woman/female').

At times, however, complex terms with a different structure are encountered. The Abipón term *eergRaik*, which is interestingly also used for 'star,' is derived from *eerg-* 'to burn, sparkle,' and Central Yup'ik *unuggsuun ~ unugcuun* is analyzable as /unuk-cuun/ 'night-device.for' and generally means 'night-light,' but 'moon' in the dialect of Nelson island. Kapingamarangi *malama* is derived from *lama* meaning 'torch' (alongside 'dry coconut leaves') and indeed also means 'lantern;' a similar situation is encountered in Hawaiian (note also that one of the terms used for 'moon' in Miskito, *ingni*, also means 'light' and 'lamp,' and that the Yuki term is semianalyzable with the identifiable constituent meaning 'light' inter alia). Furthermore, Greek *selm̄ī* is diachronically related to a verb meaning 'to shine,' and Wintu colexifies 'moon' with "something shining, bright, white, albino, rare" (Fijian *vula* colexifies 'partial albino,' the Hani term for the moon contains an element meaning 'white,' the Yuki term an element meaning "light, clear, clean, shine," and the association with 'brightness' and 'light' is also present in a group of Indo-European terms, Buck 1949: 53). Interestingly, there are also two sampled languages in which the word for 'month' is monomorphemic, and that for 'moon' (at least synchronically, since semantic shift is of course possible) secondary to it (Yay *roŋ*^{5} *duan*^{1} 'bright month' and San Mateo del Mar Huave *müm caaw* 'mother month').

As far as colexification is concerned, a ubiquitous pattern is indeed that with 'month,' occurring in 98 languages of the sample and in all areas of the world (and reconstructed for the Proto-Indo-European level as well as present in many daughter languages, Buck 1949: 54), namely Buli, Efik, Hausa, Katcha, Khoekhoe, Koyraboro Senni, Mbum, Ngambay, Noni, Dongolese Nubian, Rendille, Swahili, Yoruba, Berik, Buin, Gurindji, Kwoma, Kyaka, Lavukaleve, Mali, Muna, Ngaanyatjarra, Nunggubuyu, Meyah, Rotokas, Toaripi, Sahu, Sko, Yir Yoront, Abzakh Adyghe, Badaga, Chukchi, Japanese, Khalkha, Laz, Kolyma Yukaghir, Acoma, Biloxi, Cahuilla, Upper Chehalis, Cheyenne, Chickasaw, Highland Chontal (colexifying also 'goddess'), Ineseño Chumash, Comanche, Haida, Kiliwa, Kiowa, Lake Miwok, Nez Perce, Nuuchahnulth, Oneida, Santiago Mexquititlan Otomí, Pawnee, Quileute, Xicotepec de Juárez Totonac, Wintu, Yaqui, Yuki, Central Yup'ik, San Lucas Quiaviní Zapotec (colexifying also 'comb' and "comb for pushing back woven material on a

loom"), Copainalá Zoque, Arabela, Aymara, Carib, Cashinahua, Cavineña, Chayahuita, Cubeo, Guaraní, Huambisa, Hupda, Jarawara, Kaingang, Lengua, Macaguán, Miskito, Piro, Ancash Quechua, Rama, Sáliba, Tehuelche, Toba (by the term *ca'agoxoic,* for which compare *ca'agaxaic* 'always hurries/worries,' the original gloss being 'se apura siempre'), Tsafiki, Fijian, Great Andamanese, Hani, Hawaiian, Kapingamarangi, Bwe Karen (colexifying also 'be green, blue' inter alia), Lenakel, Malagasy, White Hmong, Rotuman, Samoan, Sedang, Takia, Tetun, and Bislama (obsoletely). This, together, with the colexification of 'sun' and 'moon' mentioned above, is chiefly responsible for the high percentage of terms for 'moon' with another colexified meaning. Buli, Bislama, Hawaiian, and Mandarin colexify 'moon' with 'moonlight,' and in Ngambay, Kyaka, and Samoan, alongside 'month,' the meaning 'season' is also colexified (in Kyaka also 'weather').

A pattern conspiciously found in two languages of Australia, Burarra and Gurindji, is that the word for 'moon' also is used to refer to a 'snail' (for Burarra, it is explicitly stated that this snail is white and moon-like in appearance). Similarly, in Nunggubuyu, the word for 'moon' also denotes 'small white grubs' and the 'chambered nautilus' (and in another Australian language, Yir Yoront, the relevant term is also glossed as "moonlike in colour and shape"); perhaps similarly, Fijian *vula* is also the name of a kind of bêche-de-mer, inter alia, and Rotuman *hulạ* is also the name of a kind of sea slug – given that the terms seem either cognate or related by borrowing into Rotuman, it is even possible that this secondary meaning is inherited or contact-related. Itzaj and Bislama colexify 'menstrual period,' and Itzaj by another term also 'bead' and 'crumble.'

Other associations include: Buli *chiik* is also the name of a "moon-amulet," and traditionally the 'soul' inter alia. Efik *ö'fiöñ* also denotes a "circular figure in painting" and is presumably related to the name of two days of the week. Buin *eekio* is also used to refer to "a period of ten days." Kyaka *kana* also means 'rock, stone' as well as 'money, coin.' Sko *ké* is also the name of a "post for hanging things" and means "catch, get, take, fetch" as a verb. Sahu colexifies "something round" generally. Toaripi *papare* contains *papa* 'grandparent, ancestor,' and Khalkha *sumija* also bears the meaning 'Monday,' which was however obsolete at the time the consulted source was written. In Kiowa, the relevant term *p̂ɑ'* can also refer to a 'river' and is the name of a game, Lake Miwok *koméenawa* is analyzable as /kóme-nawa/ 'vagina-old.man,' Nuuchahnulth colexifies 'moon' with 'thimbleberry,' and Central Yup'ik *tanqik* only means 'moon' in the dialect of Nunivak island, and denotes 'brightness' generally in other dialects. Cayapa colexifies 'moon' with 'lowered, fallen,' Tsafiki with 'excrement,' Yanomámi with a mushroom species, and Hawaiian with a "crescent-shaped fishhook" and an "eye of the snail at the end of its horn" inter alia. Rotuman *hulạ* also means 'seed, pip of fruit' and other things, and Samoan colexifies 'moon' with 'chief.'

39. The Mountain
Representation: 91%
Motivated: 53.7%
Thereof Analyzable: 11.4% Thereof Colexifying: 42.7%
Thereof by Contiguity: 15.9% Thereof by Similarity: 28.6%

Recurrent associated meanings: hill, forest, stone/rock, summit/peak,
high/top, land/earth, scarp/slope, pile, mountain range, head, valley

Cross-linguistically, lexico-semantic associations on a global scale closely mirror those mentioned by Buck (1949: 23-24) for Indo-European. Analyzable terms are most frequently based on notions such as 'high' or 'top,' alongside colexification in Yoruba. This is the case in five sampled languages: Kaluli has *hen misiyo:* 'land high.place/high.ground,' Muna *ka-bhawo* 'ABSTR-high,' Miskito *il tara* 'hill high,' Rama *king-úp* 'head/top-eye' (there are further semianalyzable terms involving *úp*, for *taisup* compare *taik* 'piece, thing, top, nose, penis, and for *kaisup* compare *kais* 'tick'?), and Fijian *dela-ni-vanua* 'top-POSS-land' and *uluni-vanua* 'head/top-POSS-land' (note also that White Hmong *roob* is probably a loanword from Chinese *chóng* 'high, lofty'); Abzakh Adyghe has *qᵒʔəśḥe* /(-)qᵒʔ(e)-śḥe/ 'arm/branch/pointed.object-up.there.' This is also the only clear instance of analyzable terms involving a body-part metaphor. Analogously to the analyzable term in Rama, 'mountain' and 'head' are colexified in Gurindji and Kaingang, where the relevant term also means 'cockscomb.'

Terms for 'mountain' often also mean 'forest' by spatial contiguity, as already discussed in section 26. This is the case in Kwoma, Basque (colexifying 'woodland' more specifically), Khalkha, Laz, Nivkh, Ineseño Chumash, Aguaruna, Huambisa, and Yay (disregarding terms glossed as 'selva' and/or 'monte' in Spanish, since these also capable of referring to 'wilderness' more generally and thus the sense colexified is unclear due to this bias exerted by the metalanguage). Similarly, Sora colexifies 'hill' with 'forest,' Cubeo has *jocu-burumu* 'wood-CLASS.VEGETATION' (there is another semianalyzable term involving a classifier for vegetation that has not been cut down in Cubeo), and Embera uses the same term for 'mountain' and, associated with different genders, various kinds of plants as well as 'plantation' and 'vegetation' generally. Also by spatial contiguity, Buin and Tsafiki colexify 'mountain' with 'valley.'

Alternatively, terms for 'mountain' may contain constituent elements meaning 'hill' and an additional element frequently conveying an additional semantic component of big size or height: San Mateo del Mar Huave has *ti-tiüc* which appears to be analyzable as 'AUG-hill,' Cheyenne *ho'honáe-vose* 'rocky-hill,' Tuscarora *yunętherʔúyʔ*, containing *-nęth(e)r-* 'hill' and *-iyu-* 'be great, be beautiful,' and Miskito *il tara* 'hill high.' There are semianalyzable terms in Bororo and Lengua. As also noted by Buck (1949: 23) with regard to 'hill,' there are sometimes "fluctuating discriminations" between 'mountain' and 'hill,' and this is reflected insofar as that these meanings are also frequently colexified. This is the case in Buli, Ngambay, Dongolese Nubian, Yoruba, Burarra, Gurindji, Kwoma, Kyaka, Muna, Ngaanyatjarara, Tasmanian (Northerastern, Middle-Eastern, and Southeastern), Toaripi, Sentani, Yir Yoront, Abzakh Adyghe, Badaga (where there is a verb of the same phonological structure meaning 'to cut'), Ket, Upper Chehalis, Cheyenne, Chickasaw, Highland Chontal, Itzaj, Kiliwa, Lake Miwok, Lesser Antillean Creole French, Santiago Mexquititlan Otomí, Pawnee, Tuscarora, Wappo, Yaqui, Yuki, San Lucas Quiaviní Zapotec, Bora, Bororo, Embera (where the meanings are associated with different genders), Hupda (also colexifying 'cliff'), Jarawara, Lengua (colexifying 'large hill' specifically), Macaguán,

Maxakalí (colexifying 'steep hill' specifically), Ancash Quechua, Rama, Tsafiki, Wayampi (dialectally), Yanomámi, Bislama, Hawaiian (colexifying 'high hill' specifically), Kapingamarangi, Malagasy, Rotuman (colexifying also "mound of earth or sand" inter alia), Samoan, Takia, and Tetun; note also that for Toaripi *raepa*, glossed as 'hill, mountain,' the lexicographer remarks that "[f]or the latter meaning the adj[ectives] rovaea (=big) or koa (=high) is often added."

As already seen in some terms, complex terms for 'mountain' at times also feature a constituent with the meaning 'land' or 'earth' generally. Further instances of this are found in Efik (*akamba obüt* 'great land/earth'), Kosarek Yale (*mok-soo* 'place-earth;' this term also means 'world' and 'land' by itself), Guaraní (*yvy-ty* 'earth-pile'), Wayampi (*i̵wi-ti̵* 'ground-big'), and possibly Cheyenne (*o'omenó* may contain *-o'ome* 'region' and *-nó* 'place'). There is a further semianalyzable term where the identifiable constituent is 'land' in Kaluli.

Buli, Abzakh Adyghe, Jarawara, and Tetun colexify 'mountain' with 'scarp, slope,' and Ngaanyatjarra, Yaqui, and Mandarin with 'mountain range.' There is, as also noted by Buck (1949) for Indo-European, a recurrent association between the meanings 'mountain' and 'stone, rock,' in the languages of the sample in Buli, Ngambay, Ngaanyatjarra (these two languages also colexify 'pebble'), Kolyma Yukaghir, Biloxi (where the colexified meaning is "round-topped hill" more specifically, the relevant term may contain an element meaning 'sharp'), Cahuilla, Oneida (also colexifying 'outcropping'), Hupda (colexifying also 'sky'), and Hawaiian (where "kind of hard stone from which adzes were made" is in addition colexified). White Hmong has *pob-tsuas* 'ball-rock.mass,' and there is a semianalyzable term featuring a consitutent meaning 'stone' in Copainalá Zoque. In Kyaka, *kyau* also means 'pile' and 'swelling,' Highland Chontal *łijuala* also means 'pile,' and Lesser Antillean Creole French *mòn* may also figuratively refer to 'a large heap' (compare Lenakel *touar* 'mountain' and *tou* 'to heap up food, yam heap'?), while Kiowa *k̂oup* may also refer to a 'knob' (again, compare Buck 1949: 23); there is also a verb of the same form meaning 'to lay several.' At times, the relevant terms also denote the 'summit' of a mountain or 'peak' more specifically, a case of meronymy. This is found in Baruya, Burarra ("top of rocky outcrop" more specifically), Abzakh Adyghe, Badaga, Khalkha ('flat mountain top' more specifically), Wintu, and Ancash Quechua; note also Abzakh Adyghe *ʔᵊa-śḥe* 'summit-up.there.'

Other associations include: Dongolese Nubian *ǧébel* also means 'desert' (a meaning which is also colexified in Sudanian Arabic, from which the term is borrowed). Buin *kumpa* is also the name of an 'edible mountain fern' and a male name (the relevant term is, however, a poetic epithet for 'mountains' only), Rendille *hál* also means 'pack camel' and also denotes a particular mountain, Kwoma *kwow* also means 'netbag' and 'womb,' and One *ala nala ~ ala nela* appears to contain *nala ~ nela* 'tooth' (there is a variant *ala palla*, for which compare *palla* 'body hair'). Sko *pì* also means 'full,' and Abzakh Adyghe *bǧə* also 'waist' and 'backrest.' Adjectivally, Badaga *male* also means "puffed up, haughty, self-important, towering above," and Basque colexifies 'mountain' with "country, wild country." Bezhta has an unusual pattern of colexification in that *mā̄* is also the default term for the 'nose' (note that Abzakh Adyghe *čape* might contain *p(e)*, meaning 'nose, beak, point' inter alia).

Khalkha colexifies 'mountain' inter alia with 'plateau' as well as "shelf, hanging rack," Ineseño Chumash with 'north,' Itzaj with 'pyramid,' Santiago Mexquititlan Otomí with 'dark,' and Tuscarora with 'billow.' For Lesser Antillean Creole French *montani*, compare *montan* 'rising, ascending,' and for Yana *ʒiigal(la) ~ ʒiigal(xi)* (which are dialectal variants), compare *ʒii-* 'smoky, foggy.' Copainalá Zoque colexifies 'mountain' with 'field,' and the Abipón term *lkaagRanRat* appears to be analyzable as *l-kaag-Ran-Rat* 'POSS.INDEF/3SG-split-CAUS-CAUS' (original Spanish gloss of the root *aag- ~ -aage ~ -aak ~ -kaak ~ -kaag* is 'morder, hender, partir, cavar'). Toba *qasoxonaxa* is also the name of a mighty being causing lightning, and is also used to refer to the 'elephant.' Hawaiian has *kuahiwi*, containing *hiwi* 'sharp ridge of mountain' and *kua*, meaning 'back, rear' inter alia. Kapingamarangi *gono duu* is analyzable as 'form/surface/hue stand/stop/belt.' Bwe Karen colexifies 'skin, shell' inter alia, and Lenakel "elevated place, village, township." Rotuman *solo*, as a verb, also means 'for the sun to sink,' Sedang colexifies 'mountain' with 'stump,' and Tetun with "country(-side)."

40. The Mushroom

Representation: 69%
Motivated: 17.8%
Thereof Analyzable: 15.4% Thereof Colexifying: 2.9%
Thereof by Contiguity: 1.0% Thereof by Similarity: 15.7%
Recurrent associated meanings: ear, faeces, spirit/devil, horse

Terms for 'mushroom,' if they are motivated, have a very marked lexico-semantic profile: motivation is very often realized by analyzable terms, and the semantic relation underlying them is very often one of similarity. Although there is wide variation as to the specific semantic source concepts to be discussed below, there are two recurring strategies that are both remarkable. One conceptualization is based on 'faeces' (see also § 6.2.3.3.) and the other one is based on 'ear.' Thus, Rendille has *u'dú-yeyyah* 'moon-faeces,' San Mateo del Mar Huave *aonts potwit* 'excrete black.vulture,' Toba *huaqajñi l-'atec* 'star 3SG.POSS-excrement' (colexifying 'shooting star'), Hawaiian *kūkae-lio* 'excrement/dung-horse,' (this term also denotes grasses where horses are pastured. For the association with 'horse,' compare Toba *cainaton* /caaỹo-naton/ 'horse-glans;' there are several variants of this term, and in one of them, the element denoting 'glans' is colexified with 'sombrero'), and Cashinahua resorts to direct colexification.

Carrier has *impiñ-dzo* 'pigeon-ear,' Central Yup'ik (Nunivak Island dialect) *tuunram ciutii* contains elements meaning 'spirit, devil' and 'ear,' Fijian has *daliga ni kalou* 'ear POSS spirit,' and Bislama *sora blong devel* 'ear POSS spirit.' Moreover, Aguaruna, Rotuman, and Samoan colexify 'ear' with (types of) 'mushroom' inter alia, and there is a semianalyzable term in Anggor; the Rotuman term *faliga* also means 'pectoral fin,' and there is also a complex term *faliga ne 'atua*, with *'atua* meaning 'ghost.'

The Samoan evidence, where the relevant term is said to denote several species of fungus, points to a potential problem: it is not possible to be sure that terms in the consulted sources really correspond to the life-form level and do not rather denote a specific

type of mushroom on the generic level. However, it seems unlikely that the cross-linguistically robust occurrence of the associations can be due only to such ambiguities due to dictionary information. For the area of Oceania, genealogical inheritance, areal spread, or a low diversity of biological diversity on the small islands of Oceania cannot be excluded as possible factors, but even then, the association is also found in areas of the world in which such factors seem unlikely. If the association with 'ear' is thus indeed a genuinely recurring phenomenon, what is the cause? Tree-growing fungi are quite widespread globally, and in some areas may even be the only type of mushroom found. Often, these fungi have a decidedly ear-like shape, as opposed to the typically cap-like appearance of soil-growing mushrooms. The Jew's ear (*Auricularia auricula-judae*), particularly mentioned in the dictionary gloss for Samoan, is an instance of this.

There are, alongside those occurring in association with the metaphorical transfer of 'ear,' also other terms making reference to spirits: Wichí has *ahot-lhu* 'spirit/soul-penis' for 'black mushroom,' and Fijian furthermore *iviu ni tēvoro* 'fan.palm POSS demon.'

There are further metaphorical patterns including associations with certain animals in languages of Africa: Hausa has *nama-n kaza* 'meat-GEN hen' (*kaza* also denotes an ungrateful person) and Kanuri *tə́mbàl kókó-be* 'drum frog-of' (there is another term in Kanuri which appears to colexify 'rainy season'); moreover, Wichí has *mawu-tonek* 'fox-liver' for 'orange mushroom.' Still other metaphor-based terms are Japanese *ki-no-ko* 'tree-GEN-child,' Kolyma Yukaghir *ara:n-paj* 'naked-woman' for a 'mushroom growing on earth,' Upper Chehalis *ɬúm=lwltx*^{w} 'wrinkle/shrink=house/building/place.where.animal.lives' and Kiliwa *phitnsmay* with the literal meaning 'little lost fart' (compare also Yay *rat*^{3} *raap*^{2} 'mushroom' with *taw*^{3} *rat*^{3} 'to break wind?). There is also an association with 'hat' or 'cap' in one language, Haida. There are several complex terms involving *dajing* 'hat, cap' one of them *kagann dajing*, with *kagann* meaning 'mouse' ("[m]ice were the physical form assumed by witches' evil spirit").

Other associations are: Efik *udïp' ek'pe* seems analyzable as 'NMLZ-hide bunch' (with *ek'pe* also meaning 'panther, leopard' inter alia). Ngambay *bbè* also means 'quiver,' and Yoruba colexifies 'chief among persons' and 'queen of ants' with 'mushroom.' Kaluli *ko:lo:* is also a "word to signal or point back to something just talked about or something just mentioned," Sahu colexifies 'mushroom' with 'rust,' Basque *ziza* also means 'lisping' as well as 'to excise tax,' and Lesser Antillean Creole French *chanpiyon* also 'champion' (due to collapse of Fr. *champignon* and *champion*). Wintu *ʔaλ* also means to 'look on, observe, watch' inter alia, *qun* in the same language also means 'mold' and 'blue.' Central Yup'ik *palurutaq* ~ *paluqutaq* (Hooper Bay and Chevak dialect) also means 'quonset hut' and 'turtle.' Cubeo *chĩchi* colexifies 'scale,' and Miskito *yula* is also used with the meanings 'dog,' 'insects,' 'little animal,' and 'parasite' inter alia, and *srapka*, another term, seems to contain *srap* 'algae, moss.' Piro colexifies 'mushroom' with 'lichen,' and Malagasy *hòlatra* also means 'scar.' In addition, the Yay term *rat*^{3} *raap*^{2} appears to contain *raap*^{2} "to carry on the two ends of a shoulder pole."

41. The Nest

Representation: 89%
Motivated: 40.0%
Thereof Analyzable: 21.0% Thereof Colexifiying: 19.0%
Thereof by Contiguity: 5.1% Thereof by Similarity: 32.1%
Recurrent associated meanings: house/home, bird, den/lair, container/box, beehive, bed, web, basket, raft, shelter, hole, grass, rubbish, egg

There is one metaphorical pattern that is more frequent than any other lexico-semantic associations found for it in the database. This is the likening of the bird's nest to the house of humans. The association can be realized by both colexification and morphologically complex terms, although the latter strategy is more common. Complex items can be most often be translated literally as 'bird house' (such as Kaluli *o:ba: a* 'bird house,' this is also the case in Efik, Kanuri, Mbum, Ngambay, Noni, Yoruba, Quileute, Miskito, and Malagasy), but some variation is found. For instance, Embera has *ĩmbaná dhe* 'bird home,' and Kiowa *tou-sǫų'n* consists of *tou* 'house' and *sǫų'n* 'grass.' The same structure is also found in Cayapa, compare also Toaripi *ori roro* 'bird rubbish/refuse/weeds' and Pawnee *rahkisiitu?* /raar-kisiit-u?/ 'ITER-flat.reed-NOM.' In turn, for the association with rubbish, compare Pipil *-tapahsul*, consisting of the 'unspecified object' prefix *ta-* and *-ku:pahsul* 'rubbish' (there is an alternative possible source for the sequence *pahsul*: *mu-pahsulua* "for one's hair to be messed up"). In San Mateo del Mar Huave *omb-iüm* is analyzable as 'hole-house' (this complex term can also refer to a 'house' itself, compare colexification of 'nest' and 'hole' in Kolyma Yukaghir). Kyaka features in addition the term *anda pingi* 'house/nest root,' Kapingamarangi *hale ngogo* 'house egg,' and Sora has *əsu:ŋtidən* /ə-'su:ŋ-'tid-ən/ 'POSS-hut.for.temporary.use-bird-NMLZ.' Moreover, there is a derived term in Kiliwa, and colexification with 'house' and/or 'home' is found in Hausa, Khoekhoe (where the term is formally derived from a verb meaning 'to build, construct), Kyaka (also colexifying "open valley area" and 'shed'), Muna, Badaga, Wintu, Maxakalí, and Lenakel (and in Tasmanian with 'hut, camp' more specifically).

A further metaphorical transfer is one from the meaning 'bed' rather than 'house' to the 'nest.' This is found in Samoan and Bislama by morphologically complex terms (*fa'a-moega* 'like-bed' and *bed blong pijin* 'bed of bird' respectively), and in Guaraní and Manange by colexification (similarly, Khoekhoe colexifies 'nest' with "sleeping place, resting place"). The Burarra term for 'nest' consists of the verb for 'to mound up,' *gapulawa,* nominalized by prefixation of the 'general' class marker *gun-* and also means 'clump, knot.' The same pattern is possibly also found in Wintu (however, the relevant term is only attested from one speaker). In four sampled languages, Muna, Kolyma Yukaghir, Jarawara, and Hupda, an association with 'container' or 'box' is found (in Welsh also with 'case'), in the case of Hupda by the morphologically complex term *hũtæ̃h cáʔ* 'bird box,' and in Muna by a term derived from a verb meaning 'to collect, receive' inter alia. Similarly, Buin and Fijian colexify 'nest' with specific types of baskets (Fijian also with other meanings). Efik has *e'fȏk i'nuën* 'sheath bird.'

A very interesting association is found in Nunggubuyu and Middle-Eastern Tasmanian: in both languages, the relevant terms can also refer to a 'raft' (in Nunggubuyu, to a raft made from paperbark specifically), and this may be a case of provenience contiguity, since (at least some) birds in this area build their nests in paperbark trees.

Common is also colexification of 'nest' with other habitations of animals: with 'den' or 'lair' in fifteen sampled languages (Kwoma, Rotokas, Waris, Basque, Greek, Khalkha, Kolyma Yukaghir, Nez Perce, Wintu, Cashinahua, Maxakalí, Hani, Mandarin, Sedang, where the relevant term colexifies 'garment, blanket,' and Yay), with 'hive' in Nivkh, Hawaiian, Mandarin, and Lesser Antillean Creole French, and with 'web' in Buli, Muna, and Ineseño Chumash. Badaga and Hawaiian also colexify 'shelter' generally.

Given that the function of a bird's nest is to lay eggs in it, it is surprising that lexico-semantic associations with 'egg' are quite infrequent. Alongside Kapingamarangi *hale ngogo* 'house egg' which was already mentioned, Ket has *eŋgaj* /eˀŋ-àj/ 'eggs-sack,' and Chukchi *kətcəɣjolɣen* is related to *kətcənjo-* glossed as 'sit in ambush for' in the consulted source, which also cites an older source stating its meaning to be 'to sit on eggs' (compare Itzaj *k'otb'aj* 'brood, nest, cluck with chicks,' derived from *k'ot* 'to brood').

Morphologically complex terms constituting semantic relations that occur only in one of the sampled languages include Swahili *kiota ~ kioto*, consisting of *ota* 'sprout' and a noun class prefix and Muna *kaofe ~ kaufe*, apparently derived from *ofe ~ ufe* to "squeeze cooked rice into a round shape, compress." Abipón features a term derived from a verb meaning 'to assemble,' Fijian *vakavevēde* may contain *vaka*, which alongside grammatical functions means 'be like, resemble,' *veve* 'crooked, bent' and *dē* 'to fix firmly.' Derivation by suffixation of classifiers is found in two languages of the Amazon: Chayahuita *pë'pëtë'* is derived from *pë'përin* 'carry' by the instrumental classifier *-të'*, and Cubeo *cɯribɯ* appears to be derived from *cɯrõ* 'place, site' by the classifier *-bɯ* for cylindrical or round objects. Khoekhoe *haires* contains *hai* 'tree, wood, plant, stick,' Badaga has a term for "nest among stones" containing a constituent meaning 'stone,' Upper Chehalis one meaning 'moss,' and another Upper Chehalis term contains a verb meaning 'to settle down, reside.' The Tuscarora term for 'nest' appears to contain the word for 'mother.' The relevant Arabela term contains a classifier for balls of fibres, the Guaraní one a constituent meaning 'pile,' Piro *sreta* is related by unknown means to *sure* 'leaf, sheet of paper,' Sáliba *juwõchẽ* appears to contain *juwo* 'hairs, feathers,' and the Great Andamanese term *ârrâm* is derived from *râm* 'to cover' by prefixation of a possessive marker. Semianalyzable terms including a constituent meaning 'bird' are attested in Khoekhoe, Kemtuik, and Kwoma.

Unique patterns of colexification include that with 'winnow' as well as 'to pour into vessel' inter alia in Efik, with 'iris of the eye' and other meanings, and, by a different term, with 'support' and 'husband' in Buin, with "bush camouflage, hide" in Ngaanyatjarra, with 'rudiment' and other meanings in Kyaka, with "(family) goods, possessions" in Sentani, with 'writing' in Sko, with 'cave' in Abzakh Adyghe, with 'bear' and 'womb' in Bezhta, with 'cradle' in Kashaya, with 'niche, nook' in Lesser Antillean Creole French, with 'straw' in Wintu, with 'base' in Carib, with 'diaper' in Cavineña, with 'hammock' and/or 'yawl' in Wichí, with 'shelter' and 'gathering place' and other meanings in Hawaiian, with 'string, rope' and by another term with 'blood' in Bwe Karen, with 'placen-

ta, afterbirth,' 'coccoon,' and 'handle' in Lenakel, with "body cavity, such as armpit, hollow of the knee" as well as 'tide, morning tide' in Mandarin (though the latter term has a different etymon, Pulleyblank 1991: 51), and with "shelf up high" in Sedang. There are extensions to the human sphere in Khalkha ('cell of a political party') and Mandarin (' "nest" of robbers'). Similarly, the Basque term *habia* may also be employed metaphorically with reference to the human sphere. Finally, note that Yuki *noh ~ noʔ* 'nest' and *noh- ~ noʔ-* 'to live' are identical segmentally.

42. The Plant

Representation: 59%
Motivated: 66.9%
Thereof Analyzable: 20.6% Thereof Colexifying: 46.8%
Thereof by Contiguity: 28.1% Thereof by Similarity: 0%
Recurrent associated meanings: tree, grass/weed, thing, grow, vegetable, flower/blossom, bush/shrub, vegetation, leaf, forest, seed, to plant, sprout/shoot, green, plantation, stalk, land/earth

It is common for terms for 'plant' to be associated in some way with a more specific type of plant further down the taxonomy (autohyponymy, in terms of Horn 1984). Thus, Khoekhoe, Ngambay, Rendille, Ngaanyatjarra, Yir Yoront, Sora, San Mateo del Mar Huave, Kashaya, Kiowa (where the relevant term is furthermore identical segmentally with that for 'to be dewy'), Nuuchahnulth, Quileute, Yaqui, Ancash and Imbabura Quechua, Fijian, Hawaiian, Rotuman, Samoan, Takia, Tetun, Yay, and Bislama colexify 'plant' with 'tree' (some also with further meanings discussed in section 65), and there is an overt term exhibiting the relationship in Bora (*úmé-hé-wu* 'tree-CL.tree-DIM'). Muna, Abzakh Adyghe, Badaga, Bezhta, Ket, Kildin Saami, Upper Chehalis (by a semianalyzable term containing an element meaning 'grow'), Cheyenne, Ineseño Chumash, Embera, Guaraní, Maxakalí, and Wayampi colexify 'plant' with 'grass' (similarly, Chickasaw colexifies 'wild plant' with 'weed' and Miskito 'plant' with 'medicinal herb'). There is a semianalyzable term where the identifiable meaning is 'vegetation, weeds' in White Hmong. Meyah, Badaga, Kildin Saami, San Mateo del Mar Huave, Santiago Mexquititlan Otomí, and Central Yup'ik colexify 'plant' with 'flower, blossom' (Badaga also with 'harvest' and 'offspring'). Rendille, Kyaka, Badaga, Nuuchahnulth, Embera, and Maxakalí colexify 'plant' with 'bush, shrub.' Moreover, San Lucas Quiaviní Zapotec *cwàà'n* is ambiguous between a narrow reading 'alfalfa' and the broad reading 'plant in general.'

Muna, Yaqui, Wayampi, and Hawaiian, as well as Ngambay, also colexify 'forest' (Abzakh Adyghe also 'decoction, medicine' and, probably from there on, 'cleaning, chemical product,' Bezhta also 'hay,' and Embera also 'mountain' and 'thicket,' see also Buck 1949: 521 for similar associations in Ancient Greek and Latin).

Dongolese Nubian, Abzakh Adyghe, Ket, Kildin Saami, Central Yup'ik (dialectally), Guaraní, and Hawaiian colexify 'plant' with 'vegetable,' and Dongolese Nubian, Rendille, Cheyenne, Embera, and Guaraní with 'vegetation' generally (similarly, Hawaiian colexifies

'greenery' and has the analyzable term *mea ulu* 'thing vegetation'). Embera and Kaingang colexify 'plant' with 'plantation.'

Moreover, it is common that 'plant' is colexified with specific parts of plants (automeronymy, if one likes). Thus, Hausa, Kyaka, Carrier, Nuuchahnulth, and Tuscarora colexify 'plant' with 'leaf' (Kwoma also with 'paper;' in Tuscarora, the 'leaf'-reading is archaic), and there are complex terms betraying this association in Yoruba (*ewé-ko* 'leaf-farm'), and Nivkh (*pan'-tjomr-ku* 'grow-leaf-PLURAL.SUFFIX'); furthermore, Carrier has the redundant complex term *yenhwoṭan*, containing *yen* 'earth' alongside *ṭan* 'leaf, plant.' Hausa colexifies 'plant' with 'root' (alongside 'source of river' and 'double sheet of paper' inter alia), and Kyaka has *renge pingi-pi* 'stem/basis/origin root-ASSOC.' Baruya, Upper Chehalis, Aguaruna, and Huambisa colexify 'plant' with 'seed' (Aguaruna by a term that might be semianalyzable, containing a constituent meaning 'field'), and Ngambay colexifies it with 'stalk, reed,' an association mirrored by the verbal Pawnee term *tat-kus* 'stalk-to.be.sitting.' Nuuchahnulth *ƛ̓aqapt* can also mean 'branch.'

However, there are also complex terms for 'plant' not of the two major types so far discussed. Nivkh *pan'-tjomr-ku* 'grow-leaf-PLURAL.SUFFIX' was already mentioned, and there is also a number of other languages where there is an association with 'to grow.' Sora, for one, has *ə-nʔeb-ən ~ ne:b-ən* '(POSS-)to.grow-N.SFX,' Hani *ja-ssaq* 'grow.tall-CLASS.SMALL,' Malagasy *zavamanìry*, analyzable as /zàvatra-manìry/ 'thing-to.grow,' and such terms are also found in Katcha and Japanese. Abzakh Adyghe has *śʼxʼentʔeyačʔe* /śʼxʼantʔe-ɣe-čʔ(e)/ 'green-ABSTR-grow,' with the item meaning 'to grow' also colexifying 'seed' and 'egg' inter alia (compare the Hawaiian term mentioned above). The association with 'growing' is also present by a derived term in Central Yup'ik and in Ineseño Chumash by an unclear morphological process. Upper Chehalis *sƛ̓aċáẏtm̓š* contains *ƛ̓aċ* 'grow' and *=tm̓š* 'land,' and Chayahuita has *no'pa quëran paporin-so'* 'earth from grow-3SG.SUB' (for this association, compare also the Carrier term mentioned above, as well as that there is a semianalyzable term with the meaning 'soil' as the identifiable constituent in Guaraní). Upper Chehalis has another term derived from another verb meaning 'to plant, to grow' colexifying 'seed,' while 'plant' and 'it is growing' are colexified in Acoma, and there are semianalyzable terms featuring a verb meaning 'to grow' in Upper Chehalis and Central Yup'ik; moreover, Kosarek Yale *bongodoba* contains *bongodob* 'everything which grows,' and note also the similarity between Yuki *ču·* 'plant' and *ču·h-* 'to grow' (see Buck 1949: 521 for this association in Indo-European). As also noted for Indo-European by Buck, Badaga, Kolyma Yukaghir, and Tuscarora betray an association of 'plant' with 'sprout, shoot' by colexification (Badaga also colexifies 'harvest' and 'offspring,' and Kolyma Yukaghir also 'beam'), and the association is realized by alternation of noun class in Swahili. Moreover, there are a number of terms in which 'plant' is associated lexically with 'to plant.' Yoruba has *ọ̀-gbìn* 'NMLZ-to.plant' (colexifying 'planter, farmer'), and *ohun ọ̀-gbìn* 'thing NMLZ-to.plant,' Hupda *yúm* for 'plant planted by humans' is analyzable as 'plant/sow.NMLZ,' and Fijian has *i kei* 'DERIV to.plant.' Mandarin and Vietnamese, as isolating languages, have complex terms for 'plant' featuring the constituents 'to plant' and 'thing,' and there is a semianalyzable term in Piro.

As noted throughout, there are some languages in which the relevant terms feature a constituent meaning 'thing.' Rendille, Kwoma, Ngaanyatjarra, and Yir Yoront colexify 'plant' and 'thing' directly (but 'grass' is excluded from the denotational range of the Rendille term; Kwoma also colexifies 'animal'), and a semianalyzable term of this kind is also found in Efik. While Abzakh Adyghe has a complex term for 'plant' with a consituent meaning 'green,' Dongolese Nubian directly colexifies 'green, light blue' with 'plant,' and there is a semianalyzable term in Ineseño Chumash.

Other associations include: in informal usage, Khoekhoe *haii* also means 'marihuana,' and Rendille *géey* also means 'dance, song.' Kaluli *é* also denotes a 'seedling' specifically. The Basque term *landare* is derived from *landa* 'field.' Lesser Antillean Creole French *plan* also means 'plan, scheme' (due to collapse of Fr. *plante* and *plan*), Kiowa *goup* also 'vine,' and Nuuchahnulth *ƛaqapt* also 'Kinnikinnick, Bearberry.' Guaraní *yva ra'y* is analyzable as 'fruit DIM,' and *ka'avo* appears to contain *ka'a* 'mate,' which according to another consulted source alongside 'mate' also means 'plant' by itself, as well as 'vegetation.' Bwe Karen *-mu* also means 'day' inter alia, Hawaiian *mea ulu* is analyzable as 'thing vegetation,' Rotuman *'ại* colexifies 'stiff, rigid' and other meanings, *hū* in the same language also means, inter alia, 'lower end,' and Takia *ai* also means 'pelvis.'

43. The Puddle

Representation: 46%
Motivated: 85.6%
Thereof Analyzable: 68.1% Thereof Colexifying: 19.6%
Thereof by Contiguity: 53.7% Thereof by Similarity: 26.0%
Recurrent associated meanings: pond/lake, water, swamp, stand/sit/be stagnant, mud, hole, spring/well, lagoon, rain, pit

Often, terms for 'puddle' (not distinguishing between 'puddle' and 'pool') are contiguity-based complex terms of the lexical type, with one of the constituents being a word for 'water.' There is a recurrent subtype, namely that with terms meaning 'to stand, to sit' or 'stagnant' acting as the second constituent, as in Ineseño Chumash *s-qil-ilik'in* '3SG/3SG.POSS-water-stand.' Such terms are also found in Efik, Abzakh Adyghe, Pawnee, and Tetun (where 'to stagnate' is colexified with 'to form puddles'). San Mateo del Mar Huave has *ndorrop yow* 'hole water,' and a term with this structure is also found in Lesser Antillean Creole French. Sora has *rupa:'luŋ'dan* /rupa:-'lˀu:ŋ-dˀa-n/ 'hole-pit-water-N.SFX,' Piro *tkomha* /tkome-ha/ 'small.hole-water/eye,' and the meanings are directly colexified in Greek; similarly, Yoruba has *kòtò ketere* 'pit small.' Aymara has *uma uma* (reduplicated from *uma* 'water'), and precisely the same structure is found in Samoan. Dadibi has *pu áí ge* /pu aí ge/ 'mud water nut/egg/small.object' for 'puddle of dirty water' specifically; note also colexification of 'puddle' with 'mud' in Basque (also with 'drop,' 'waterhole,' and 'reservior'), Khalkha, Aguaruna, Wayampi ('muddy ground' more specifically), and a semianalyzable term where the identifiable constituent is 'dirt' in Chayahuita, and one with 'earth' and another one with 'adobe' in Bora. Pipil (Cuisnahuat dialect) has *ta:l-a:-pu:ni* 'ground-water-be.born,' which colexifies 'puddle' with 'swamp' and 'spring, well.' The

former pattern of colexification is attested also in Muna, Basque, Khalkha, San Mateo del Mar Huave, Arabela, and Bora, and the latter in Basque, Oneida, and Hawaiian. Mbum has *ɓì-mbàm* 'at-rain' and Berik *aro fo* 'rain water/lake/river,' while Nunggubuyu *wargaḻg* is glossed as "rainwater on ground (including puddles)."

Other complex terms involving a constituent meaning 'water' are Efik *mkpö-diök'hö-möñ* 'thing-be.placed-water,' glossed as "water surrounded by land, collected in a depression; a pool; a pond," Kaluli *ho:n wakan*, where *ho:n* is 'water' and *wakan* the name of a "inedible wild taro-like plant that grows by water," Ket *hɤ́nul*, analyzable as /hɤna-ūl/ 'small water,' Cheyenne *tsé-a'kóom-oëha*, containing *tsé-* 'that which is' and *oëha* 'water,' Kashaya *ʔahqʰa šu-naca·la-w* 'water by.pulling-remain-ABS,' Oneida *kahnekóníʔ,* analyzable as /ka-hnek-No-ʔ/ 'NEUT.AGENT-liquid/liqour-be.in.water/cook.in.water-STAT,' Central Yup'ik *meqcarrluk*, perhaps containing *meq* 'fresh water' and *-rrluk* 'thing that has departed from its natural state,' Guaraní *y-no'õ* 'water-accumulation,' Piro *tkomha* /tkome-ha/ 'small.hole-water/eye,' Hawaiian *laha-laha wai* 'RED-extended water,' Rotuman *tạn kạlu* 'water encircle' and *tạn häe* 'water contain' (these terms also mean 'lake,' and the latter also 'bay, inlet'). There are semianalyzable terms in One, Chickasaw, Piro, Toba, and Tsafiki, as well as one featuring a constituent meaning 'liquid' in Yanomámi.

Furthermore, Efik, Baruya, Buin, Kyaka, Rotokas, Kosarek Yale, Abzakh Adyghe, Cheyenne, Haida, Lesser Antillean Creole French, Tuscarora, Copainalá Zoque, Bora, Cashinahua, Guaraní, Ancash Quechua, Hawaiian, Lenakel, Rotuman, Sedang, and Yay use a single term for both 'lake, pond' and 'puddle' (Lenakel colexifies "pool on the reef at low tide" more specifically; Lesser Antillean Creole French colexifies also 'mast, pole'). Some of them are analyzable, their internal structure being discussed in section 34; furthermore, Yoruba has *ọ̀gọ̀dọ̀ kekere* 'pond small' and there is a derived term in Hawaiian. In Copainalá Zoque and Ancash Quechua, 'lagoon' is colexified in addition.

Other associations include: Buin *rurugapau* also denotes a "flooding on road, gutter, stream," Muna *tobhi* can also refer to "the deepest part of a river or the sea" inter alia, and the variant *šalba* of Khalkha *šalbaɣa ~ šalbaɣaɣ ~ šalba* 'pool, puddle, mud' also means 'quick, quickly.' Bora has *adó-wa* 'drink.NMLZ-CL.pond' for a "well or pond in the bush where animals drink (slighlty salty) water." Embera *nãmbúa* means 'puddle, pool' with masculine gender and 'profundity, depth' with feminine gender. Yanomámi colexifies 'puddle' with 'ditch,' Great Andamanese *elâkàkōdo* might contain *kōdo*, 'coil of rope' alongside the possessive prefix *âkà-*. Hawaiian colexifies 'puddle' with "small pool for stocking fish spawn," 'cistern,' 'mollusc,' and other meanings. Another term, *hāpuna*, is figuratively used with the meaning 'child,' and, due to English influence, 'harpoon.' Finally, Sedang *tóng* also is the name of a kind of grass.

44. The Rain

Representation: 99%
Motivated: 28.9%
Thereof Analyzable: 8.8% Thereof Colexifying: 21.4%
Thereof by Contiguity: 21.5% Thereof by Similarity: 2.7%

Recurrent associated meanings: water/liquid, cloud, rainy season, storm/rainstorm, sky, rain shower, neck, day, calabash for carrying water, descend, lake/pond, weather, fall, stone

'Rain,' as one might intuitively have expected, is a meaning that is not very frequently expressed by motivated terms cross-linguistically. The most common association, by configurational contiguity, is that with 'water' (further associations due to this pattern of colexification, for instance that with 'river,' are not discussed here, though see sections 34 and 47 as well as § 6.2.2.5.). This pattern is also suggested for Indo-European in diachrony (Buck 1949: 68). Sixteen of the sampled languages, namely Hausa (which also has the optional complex term *ruwan sama* 'water/rain sky;' *ruwa* also has many other meanings, among them 'juice'), Anggor, Gurindji, Ngaanyatjarra (here also 'waterhole' is colexifed), Waris, Kosarek Yale (where the relevant term also means 'life-sap, vitality' as well as 'talk, criticism'), Cheyenne, Itzaj, Pipil (colexifying also 'well' and 'pool'), Lesser Antillean Creole French (colexifying also 'sweat'), Xicotepec de Juárez Totonac, Aguaruna, Huambisa, Hupda, Jarawara, and Miskito directly colexify 'rain' with 'water.' Alternatively, a few sampled languages also feature analyzable terms that may be either characterized by contiguity or by similarity, in both cases with 'water' acting as contiguity anchor. Examples of the former are Koyraboro Senni (Labbezanga dialect) *beene-hari* 'sky-water,' Ket *ules* /ūl-ēs/ 'water-sky,' Bororo *bu-butu* 'water-fall/descent/birth' (compare Dadibi *tulubage*, presumably analyzable as /tulubo-ge/ 'fall.down-nut/egg/small.object,' as well as colexification of 'rain' and 'to descend' in Kiowa), and examples of the latter are San Lucas Quiaviní Zapotec *nnyi'sgyihah*, analyzable as /nnyi'ihs-gyihah/ 'water-stone' (compare also Tsafiki *suhuá*, analyzable as /su-hua/ 'stone-big') and Rama *yát si* 'abscess water.' Further semianalyzable terms with 'water' are found in Bororo and Rama, and similarly, Miskito has *pura laya* 'above liquid' for 'rainwater' and another semianalyzable term featuring *laya,* while Lenakel *nihi-n* may possibly, according to the source, be analyzable as 'liquid-3SG.POSS.' Four languages in the sample, Mbum, Nunggubuyu, Yir Yoront, and Cahuilla, colexify 'rain' and 'cloud(s)' or more specifically 'raincloud' by provenience contiguity (Yir Yoront also uses this term for the 'rainbow serpent,' compare section 44), while in Kiliwa, 'rain' is *kwiy h+uhaa-k* 'cloud 3+arrive-HR,' and in Khoekhoe, the same root yields terms for 'rain, thunderstorm' and 'cloud, raincloud,' with nominal designants distinguishing the readings. In three sampled languages, Badaga, Pawnee, and Abipón, 'rain' is colexified with 'storm' or 'rainstorm' (similarly, Southeastern Tasmanian colexifies 'rain' and 'thunderstorm'), and in another three, Badaga, Arabela, and Yanomámi, relevant terms may also refer to a 'rain shower' (as is the case in Irish, Buck 1949: 68). Nivkh colexifies 'rain' with 'weather,' for which compare dialectal Central Yup'ik *cella-lluk* 'world/outdoors/weather-bad.'

In four other languages, Gurindji, Muna, Wayampi, and Hani 'rain' and 'rainy season' are colexified (in Hani also 'summer'). Bezhta and Sentani colexify 'rain' with 'day' (according to Nikolayev and Starostin 1994, this pattern is due to accidental phonological changes in Bezhta). Furthermore, two languages of South America colexifying 'rain' and 'water,' Aguaruna and Huambisa, also colexify 'calabash used to carry water' in the same

term (Aguaruna in addition, 'juice' and 'mucous membrane'). Katcha *thimbidɔ ~ thɔmbɔdɔ* also means 'sky, heaven,' and Manange also colexifies 'rain' with 'sky' (note that according to Dixon 1982: 69, in dialects of Dyirbal a cognate means 'sky' in northern dialects and 'rain' in related languages, so this pattern may be more frequent cross-linguistically). Itzaj and Jarawara colexify 'rain' with 'lake.' Kapingamarangi and Samoan, presumably by homonymy, colexify 'rain' with 'neck.'

Other associations include: the Buli term *ngmoruk* also rarely refers to 'thunder, lightning,' while *wen-zuk*, perhaps analyzable as 'sky-head' means 'up, upwards, above' normally, but is used for 'rain' by the section of the clan that is responsible for rain and which hence must not utter the ordinary term. For Koyraboro Senni *baana* compare *baan-a*, meaning 'to be soft'? Rendille colexifies 'rain' with 'God,' while Burarra *yorr* is also the name of a type of shellfish and Meyah *mós* also means 'fish' generally. Kosarek Yale *mok*, dialectally meaning 'rain,' also without dialect restrictions means "place, area" and "side sprout, offshoot." Abzakh Adyghe *šχ(e)* can also mean 'to plaster, to fuse,' Chukchi *il(ə)il* is connected (reduplication?) to *ilə* 'damp,' Greek *vrochḯ* is connected to *vréchō* 'to dampen,' Japanese *ame* with different prosodic properties also denotes 'candy, sweets,' and Nivkh *lyx* also means 'rainy' and 'wet.' Nuuchahnulth *ƛ̓iic* 'light rain' also denotes "raining mist, spray," while Yana *bariƛ̓u ~ ~ bareeƛ̓u* contains *ba-* 'to spill, to flow' and *-ri(ƛ̓u)* "down, downhill, on the ground." Bororo *butaodoge* is also the name for spirits causing rain. Embera *kúe* means 'rain' with feminine gender and 'heavy downpour' with neuter gender. The denotational range of the Jarawara term *isi/iso* includes 'leg, lower leg,' 'handle,' 'stalk,' 'rain,' 'hasta,' and 'seedless fruit," while Macaguán *em* also means 'winter.' As a verb, Piro *hina* also means 'to come,' Sáliba *oxo* also means 'leaves,' and Wichí *iwumcho'* inter alia contains *wu* 'make' and the locative suffix *-cho'* 'under.' Bwe Karen colexifies 'bug' and other things, while Lesser Antillean Creole French *lapli* is also used figuratively with the meaning 'shower.' Malagasy *òrana* also denotes the 'crayfish,' and Mandarin *yu3* also means 'and.'

45. The Rainbow

Representation: 90%
Motivated: 33.7%
Thereof Analyzable: 28.8% Thereof Colexifying: 6.3%
Thereof by Contiguity: 3.4% Thereof by Similarity: 30.8%
Recurrent associated meanings: snake, bow/arc/bend, sky, god, rain, snare, color, thunder, rope, cloud

The 'rainbow' is a concept predominantly expressed by metaphor-driven terms. At the same time, it is also a meaning for which these conceptualizations very frequently exhibit what seem to be areal patterns, both on a large and on a small scale. The two most common associations are on the one hand that with terms meaning 'arc' or 'bow' or terms meaning 'bend' or 'bent,' which is most common in the Old World, and on the other hand that with 'snake,' which is most common in the Americas and New Guinea. The former pattern is attested in Greek (*ouránio tóxo* 'heavenly/relating.to.sky bow/arc'), Kildin Saami

(*tīr'mes'-jūkks* 'thunder-bow'), Welsh (*bwa'r Drindod*, as well as *bwa'r arch*, both containing *bwa* 'bow' and the latter term *arch* 'ark'), Kiliwa (*s-?+nwaaw=x-u?+siiw=h+qhaa-tay* 'INST.LONG-DN+bend=CAUS-OBL+??=3+shoot-FREQ), Yuki (*si·k wos* 'blue/green bent;' the analysis is considered questionable by lexicographers, there also is the alternative term *sikwástlik* /si·k-wástlik/ 'blue/green-stand'), Kaingang (*ta vyj* 'rain bow,' alongside *ta no* 'rain arrow'), and Tetun (*arku iha lalehan* 'arc have sky' and *arku-iris* 'arc-sky,' both apparently calqued from Portuguese). Moreover, Rendille colexifies 'rainbow' with "arc(s) of stones" for rituals and "barrier of stones," Swahili *upinde* consists of *pinda* 'bend' and a noun class prefix, and there are semianalyzable terms in Badaga and Quileute (where the relevant term probably contains an element meaning 'bent'). Interesting to note is also that Buli *nagortom*, a loanword from Twi *nyankopon-ton*, is folk-etymologized to *Naawen gori tom* "god has made a bow." In fact, associations between 'rainbow' and 'thunder,' as betrayed in Kildin Saami, are an identifiable areal pattern of Eurasia themselves, also occurring in the sample in Ket *aqqot* /ekŋ-qoˀt/ 'thunder-path' (compare also Cashinahua *navan bai tapia*, containing *nava* 'dance, singing,' *bai* 'path/river,' and *tapi* 'firefly') and Nivkh *lyj petr* 'thunder ornament' (*petr*, more specifically, is the name for a multicoloured ornament worn on shoes); the term also means 'ulcer,' 'sore,' or 'wound' (by virtue of them changing colors when healing?). The phenomenon is discussed in Räsänen (1947), see also § 6.4.3.5.

The other major association is that with 'snake' (or a specific snake species), occurring by colexification in Burarra (Gun-nartpa dialect), Nunggubuyu, Toaripi (where *lavai* is at the same time the name of a particular snake species as well as 'tortoise' and 'dolphin' inter alia), Yaqui, Bororo, Jarawara (also colexifying 'jungle monster'), and by analyzable terms in One (*suwol ilwola* 'snake shadow'), Kosarek Yale (*mano yame* 'snake soul/image,' this term is said to also denote the 'spirit of the snake'), Kashaya (*muša·laqol*, analyzable as /muša·la-ʔahqol/ 'snake-tall'), Aguaruna (*págki wajáu* 'boa resting'), and Rama (*shírking núngkit* 'boa throat'); semianalyzable terms exist in Kaluli, Kyaka (where 'snake' is colexified with 'grub' and other like creatures), and Chayahuita. For Kwoma, the source notes that "[t]he rainbow is often identified as saliva spat out by a snake ..." There is a well-known mythological complex in cultures of Australia revolving around the Rainbow Serpent (see contributions in Buchler and Maddocks 1978), and this is evidenced by colexification of 'rainbow' with 'snake' in Nunggubuyu and with specific snake species in Burarra (and note colexification of 'rain,' 'raincloud,' and the 'rainbow snake' in Yir Yoront; there are semianalyzable terms where the identifiable constituent is 'rain' in Khoekhoe and Bislama). Mead (1933) shows that beliefs of the Rainbow Serpent also occur in New Guinea, and points to several similarities between the myths of the Arapesh and those typical of Australia, and Brumbaugh (1987) reports on the Rainbow Serpent as represented in the mythology of the Feranmin and other Mountain Ok groups, also noting that it "in behavior and attributes ... corresponds closely" (Brumbaugh 1987: 32) to the ethnographic evidence from Australia. Neither of these authors makes a very strong case for continuity between the beliefs in New Guinea and Australia, although this position appears to be implicit at least in the title and discussion of Mead (1933). A remarkable parallel is Toba, where *quemoxonalo* ~ *qamoxonalo* 'rainbow' (containing *nquemoxon* 'grasps violently') also denotes a mythological great viper which punishes by causing an earth-

quake if a menstruating woman looks for water (the Marsalai, the Rainbow Snake in Arapesh mythology, pursues a menstruating woman in a story reported by Mead 1933: 41, and there are taboos forbidding menstruating women to frequent places associated with Marsalai according to Mead 1933: 43, and in Feranmin mythology the movements of the Magalim in the earth causes earthquakes, according to Brumbaugh 1987: 27).

A pattern that is more clearly due to areal spread is the association between 'rainbow' and 'snare' in North America. In the sample, this is found by colexification in Cheyenne (here, the relevant term also conveys the meanings 'fishhook,' 'fishing line,' and 'fishing pole') and Lakhota, and by the complex term *hoḳwez-pił* 'cold-snare' in Carrier. For Lakhota, the consulted source remarks: "The Indians believed that the rainbow caused the end of a rainstorm by trapping it, so that no more rain could get through" and Hall (1997: 56) states that " 'snare' or 'trap' was a common Plains name for the rainbow because rain disappeared when the rainbow appeared." Note also that in Tuscarora, the same root *-wenę-* yields terms for both 'rainbow' and 'iron.'

There are, of course, also lexico-semantic associations which do not betray any clear areality in their distribution. These include the association with 'color' in Kiowa (*tsoue-kuɑt* 'water-painted,' literal translation by lexicographer: "many-colored") and Yanomámi (*õmayari no mayõ* 'evil.spirit color footprint'). The association is formally realized by colexification in Wintu, where *sa·q* means 'colored,' 'colors,' and 'rainbow' and is in turn probably related to *sa·q* 'to bleed' (Wintu in addition features complex term on the basis of *sa·q*, and note also the redundant Mandarin term *cai3-hong2* 'color-rainbow'), that with 'rope' in Baruya (*byaangwɨla*, literally 'light rope') and Blackfoot (*náápiwa otó'piim ~ náápiwa otokáa'tsis* 'Naapi's rope,' Naapi being the name of the trickster and creator god), that with 'cloud' in Chukchi (*celgia-jaik* 'red-cloud') and Hawaiian (*ao akua* 'cloud god,' compare also the Rendille children's term *irtiyyó=hí Waahk* 'beads=POSS god,' Chickasaw *Chihoowa i̲naalhpisa'* which contains *chihoowa* 'god' and *nannalhpisa'* 'promise,' and Malagasy *antsiben'andriamànitra* /àntsi-bè-n-andriamànitra/ 'knife-big-GEN-god'), with 'moon' also in Chickasaw (*ninak ontoomi* 'moon shine'), and at least by a semianalyzable term in Yoruba (Chickasaw, in addition, also has other semianalyzable terms presumably containing *hashi* 'sun, moon;' one of them colexifies 'rings around the moon'). In addition, conceptualizations in which 'rain' and 'sky' act as contiguity anchors are frequent. For 'rain,' these include Khoekhoe *tū-ǃhana-b* 'rain-garden-3SG.MASC' (which is restricted to Bible translations) and *ǃgao-ǃhana-b* 'stop.raining-garden-3SG.MASC,' Mbum *mbàm-pélé* 'rain-tomorrow,' Abipón *oah-eta* 'rain-AGT,' and the Kaingang terms already mentioned above (furthermore, seminanalyzable terms in Anggor and Kwoma contain an identifiable constituent meaning 'rain'). For 'sky,' they include Basque *ortzadar ~ ostadar* /ortzi-adar/ 'sky-horn,' Laz *m3a-ort'apu* 'sky-belt' (maybe the association with 'belt' is an areal pattern of the Caucasus, given that Bezhta *mašola* is perhaps borrowed from Georgian *ašuni* 'belt'), Kolyma Yukaghir *kužu:n-šöriľə* and *kužu:d-oŋora:* 'sky-tongue' (as well as *jukud-onora:* 'small tongue'), the Haida "story word" *qwii sdal* 'sky slope,' Tuscarora *yurę̆hyáhuks*, containing the roots *-rę̆hy-* 'sky' and *-huk-* 'to light up' and colexifying 'Aurora Borealis' and 'Milky Way,' Wayampi *ɨwa-lɛwa* 'sky-on,' Yanomámi *hetu kë̈kɨ shĩĩ* , consisting of *hetu* 'sky,' *shĩĩ* 'light' and the collective quantal classifier *kë̈kɨ* (see § 4.4.1.1.), Fijian *drō-drō-lagi* 'run.away-

RED-sky,' as well as the associations with 'bow' in Greek, where 'sky' acts as a contiguity anchor. Moreover, there is a semianalyzable term where the identifiable constituent is 'sky' in Lesser Antillean Creole French.

As becomes clear from the above list, the conceptualization in each language is widely different, in spite of the common semantic element acting as the contiguity anchor. Unsurprisingly, there are even more unique and culture-specific metaphorical conceptualizations for this concept. These include: Hausa *bakan gizo*, which also denotes a "single arch in a roof," contains *baka* 'mouth, bow' and *gizo* 'mythical spider,' while Dongolese Nubian *káǧibbɛ̄* is literally 'killer of horses.' Buin *iroro(na)* is also an "epithet for males in songs." The Kwoma term *wariipoy* "green of rainbow, 'liver' of rainbow" is also used for a 'type of small tree that grows near trees in forests.' The Rotokas term *govugovuto* seems to contain *govugovu* 'clean out, purge.' Khalkha *solungɣa* also denotes the 'weasel' or the 'Siberian marten.' Sora *iliŋˈboːŋən ~ ˈilimˈboːŋən* is built around the noun root *ˈboːŋ-* 'class of deities,' but the additional material remains obscure. Cahuilla *píyaxat* also denotes a 'worm with two horns.' Comanche is unique in having a term for the 'rainbow,' *pisi maʔrokóoʔ*, that is literally 'infected thumb.' The Itzaj term, colexifying "fog rising from earth" is *kis witz* 'fart hill.' For Lake Miwok *káccakaca* 'rainbow,' compare perhaps *kacáakaca* 'bluebird, *Sialia mexicana*,' a multicolored bird with blue, gray, and orange plumage. Pawnee *huraahkipic* is semianalyzable: it contains *huraar* 'be land.' Yana *laʔi-yaa* is analyzable as 'navel person;' the term means 'newborn baby' in Central Yana and 'rainbow' in Northern Yana. The Central Yup'ik term *agluryak* is derived from *agluq* 'ridgepole, center beam of a structure' by means of addition of the postbase (cf. § 4.4.2.) *-yak* 'thing similar to,' and San Lucas Quiaviní Zapotec *garre'ed bihih* is literally 'cart air/wind;' a variant of the term is *garre'ed gyeht*, probably analyzable as 'cart squash/pumpkin.' Bora *tuúhi* might contain the classifier for small rivers *-hi*, Bororo *jure* colexifies 'rainbow' with 'dance' as well as 'queue,' the Cubeo term *náme* is also used to refer to a 'string of liana,' while Ancash Quechua *turmanyay* also denotes 'gas that emanates from the earth.' Fijian *mudu* or *valemudu* denotes a half rainbow (*mudu* means 'cut off, ceased, ended,' *vale* is 'house'), Hawaiian *ānuenue* is also used to refer to "the scallop-like design on tapa and tapa-beater," and *haka ʻula a kāne*, a poetic term for the 'rainbow,' is analyzable as 'perch red POSS Kāne.' White Hmong *zaj-sawv* is literally 'dragon-rise,' and Tetun *baur* colexifies 'rainbow' with 'to cheat, swindle,' and *namerak* with "to grow murky." Sedang *kia pơtea* apparently contains *kia* 'ghost,' Vietnamese *cầu vồng* is analyzable as 'bridge curved,' while Yay *roŋ*^{5} *ʔwa*^{1} contains *roŋ*^{5} 'bright.' The San Mateo del Mar Huave term *ndequiamb poj* seems to contain *poj* 'terrestrial turtle.' Finally, Piro colexifies 'rainbow' with 'pus' and the meaning to be treated in the following section: the 'resin.'

46. The Resin

Representation: 66%
Motivated: 49.5%
Thereof Analyzable: 27.0% Thereof Colexifying: 24.6%
Thereof by Contiguity: 5.0% Thereof by Similarity: 34.2%

Recurrent associated meanings: water/liquid/juice, tree, blood, tar, stick/sticky, glue, milk, honey, wax, syrup, rust, molass, semen, cud, phlegm, birdlime, pus

Most frequent for this meaning (for which 'sap' and 'gum' were accepted as proxies if it was clear that indeed sap of plants and trees is meant) are lexical associations with 'water' and/or more generally 'liquid' or, less generally, 'juice.' One or more of these meanings are colexified in Lavukaleve, Bezhta, Ket, Sora, Maxakali, and Hawaiian (Bezhta also colexifies 'drink' and Ket also 'alcoholic beverage'). More common, however, are terms in which one of these meanings merely is one of the constituents in analyzable terms, as in Nivkh *tiɣr-čoχ* 'tree-juice.' Alongside Nivkh, such terms with 'tree' as the other constituent are also found in Kanuri, Japanese, Cheyenne, Chickasaw, Lesser Antillean Creole French, Yaqui, Carib (colexifying 'torch'), Imbabura Quechua, Mandarin, White Hmong, and the alternative where 'bark' rather than 'tree' is used as the contiguity anchor is attested in Piro (*mta-ha* 'bark water'). A semianalyzable term in which one constituent can be identified to mean 'water' is found in Kyaka, and further variations of this pattern are Bororo *bato kuru* 'mangaveira liquid' and Ancash Quechua *hacha-pa wiqi-n* 'plant-GEN juice/tear-3SG.' In Santiago Mexquititlan Otomí, the relevant term is *dehe 'yo-mu(n)hño ngizá* /dehe 'yo-mu(n)hño ngi-zaa/ 'water walk-good sap-tree.' The language also has another term for 'resin,' namely *'ba zaa* 'milk tree.' Associations with 'milk' are also found by colexification in Kwoma, Sora, and Ineseño Chumash.

An association by colexification between 'resin' and 'blood' is attested in Kwoma, Cubeo, Maxakalí, Fijian, and Hawaiian. Also note Sko *hí* 'sap' and *hì* 'blood' as well as that Jarawara colexifies 'blood' with 'red sap of certain trees' specifically and, by another term, also colexifies "puss," which is probably an error in the source for 'pus.' At any rate, 'pus' and 'sap, resin' are colexified in Sora and Piro (alongside 'rainbow' in the latter language). Bislama has *blad blong tri* 'blood of tree.' Colexification with 'honey' is found in Nez Perce, Oneida, and Bwe Karen (which also colexifies 'to be pure, clean' inter alia). Three sampled languages, Tuscarora, Hawaiian, and Tetun, colexify 'resin' with 'wax' (similarly, Rotuman colexifies 'sealing wax' specifically), Nez Perce, Oneida and Tuscarora colexify 'syrup,' Nez Perce and Tuscarora also 'molasses,' and Nez Perce also 'sorghum.' Two of these languages, Tuscarora and Hawaiian, also use the relevant term for 'tar,' an association also occurring in Efik, Aguaruna, and Bora; the Tuscarora term *uθrę̀·weh* also colexifies 'cement' as well as 'jam' and 'jelly,' and the Hawaiian one also 'printers' type' and 'sinker on a fishing line' inter alia. Nez Perce furthermore colexifies 'cud chewing' (by another term than that participating in the above patterns), and 'resin' and 'cud' are also colexified in Chukchi. Four other languages, Buli, Waris (where the relevant term *nénél* appears to contain *né* 'forest, forest product'), Fijian, and Tetun colexify 'resin' with 'glue,' either by provenience contiguity if resin is actually used as glue, or by perceptual similarity based on their common stickiness. In fact, there are languages in which the words for 'resin' explicitly make reference to this. For instance, Hawaiian *pīlali* denotes the "[h]ardened sap … of the kukui tree, gum; resin, birdlime; wax" but also means "gummy, sticky" inter alia, and the association with 'sticky' is also found in Berik and Wayampi by colexification. Welsh has *defnydd gludiog o coed* 'matter sticky of tree,' Lenakel *nouanehapwiit*, containing *noua* 'fruit' and

apwiit 'stick, cleave to,' and a semianalyzable term in which the meaning 'sticky' or 'stick' figures occurs in Kapingamarangi. Furthermore, there are a number of mostly metaphor-driven complex terms for 'resin' in which 'tree' acts as contiguity anchor. These include Ngambay *nɔn-kake* 'tear-tree,' San Mateo del Mar Huave *aonts xiül* 'excrete tree,' Guaraní *yvyra ry'ái* 'tree sweat,' Piro (*gagmuna-)kshi* '(tree-)rainbow,' Hawaiian *hū lāʻau* 'swell tree,' Malagasy *tsironkàzo*, analyzable as /tsìro-hàzo/ 'taste-tree/wood,' Sedang *chhá lóang* 'split tree/twig,' as well as Takia *ai pat-an* 'tree kidney-3SG.' There are semianalyzable terms with the identifiable constituent meaning 'tree' in Kosarek Yale, Cavineña, and Sáliba. Sahu and Hawaiian colexify 'resin' with 'birdlime.' Two languages of Mesoamerica, Itzaj and San Lucas Quiaviní Zapotec, colexify 'resin' with 'rust' (San Lucas Quiaviní Zapotec also with 'nectar' and 'lymph'). Toba and Hawaiian colexify 'resin' with 'phlegm,' and Miskito and Hawaiian with 'semen.'

Other associations include: Gurindji *tinung* also denotes the "bloodwood sap" specifically, Kosarek Yale *keles* 'resin used for glueing and tightening a drumskin,' according to the consulted source, might be related to *kel* 'female' and *es* 'flower,' and *widi*, also a name for resin used for tightening a drumskin, also denotes a variety of sugar cane. Abipón *liciRa* is derived from *-ici ~ -icir-* 'assemble, connect,' while Arabela *riya-ca* appears to be analyzable as 'star:PL-CLASS.FRUIT.' Bororo colexifies 'resin' with 'rubber,' and Cayapa also means 'cresent in a river.' Chayahuita *yaqui'* is derived from the verb *yaquirin* 'to cut well or completely' by means of suffixation of the classifier for liquids, *-i'*. Huambisa colexifies 'resin' with 'tattoo,' and Kaingang *jẽnjo* might contain *jẽn* 'to eat.' Guaraní *aysy* also means 'gluten.' Rama *shubli ~ ubli ~ yubli* can also refer to a 'stain' or a 'secret.' Hani *ziq* also means 'hemp,' a relevant Hawaiian term colexifies 'to roll, turn' inter alia, and Rotuman *pulu*, which can refer to "any adhesive substance" in general, can also be used with the meaning 'seal.' Hawaiian *pīlali* may contain *lali*, meaning 'greasy' inter alia. Finally, probably accidentally, Tetun colexifies 'resin' with 'candlenut' and 'to prune, clip,' and Yay *θa*^{3} also denotes "any of various devices having a wheel."

47. The River

Representation: 96%
Motivated: 43.6%
Thereof Analyzable: 11.0% Thereof Colexifying: 33.2%
Thereof by Contiguity: 29.2% Thereof by Similarity: 2.6%
Recurrent associated meanings: water, spring/well, lake/pond, juice, valley, river bed, ocean, way/trail/track, big, rain, flow/run, channel/drain/ditch/gully, floodwater, Milky Way

The most frequent association in this case, clearly, is with 'water.' Many languages in the sample colexify 'river' (or 'stream,' 'creek,' etc., which were accepted as proxies) with 'water,' and many of them also use the general term for 'water' also for other bodies of water, such as a 'lake' or a 'spring.' These will not be discussed here, see sections 34, 44, and 56, and especially § 6.2.2.5. for discussion. Terms which colexify 'water' (or 'fresh water' specifically) and 'river' are found in Efik (also colexifying 'tide'), Mbum, Ngambay,

Anggor, Baruya, Berik, Buin (where the relevant term is also a "general name for all ruta designs" as well as an epithet for a "plumb child"), Kyaka, Mali, Muna, Meyah, One, Sko (where also other meanings are colexified), Waris, Kosarek Yale (where the relevant term also means 'life-sap, vitality' and 'talk, criticism'), Abzakh Adyghe, Badaga (also colexifying 'floodwater,' as is the case in Ngaanyatjarra and Abzakh Adyghe), Cahuilla, Upper Chehalis, Haida, Kashaya, Pipil, Quileute, Wintu, Copainalá Zoque, Bororo, Cayapa, Chayahuita, Guaraní, Huambisa, Jarawara, Kaingang, Macaguán, Tsafiki, Wayampi, Yanomámi, Bislama, White Hmong, Sedang (where further apparently unrelated meanings are colexified), and Takia (40 languages all in all and thus a little more than twenty-five per cent of the sampled languages). Kosarek Yale, Abzakh Adyghe, Kashaya, Copainalá Zoque, Tsafiki, and Bislama colexify also 'juice,' and sometimes still further meanings.

Some of the languages just mentioned also have special dedicated terms for 'river' alongside the colexifying term, such as Efik, which has *akpa* 'river' alongside the general term *'möñ*. Also, some languages have optional analyzable terms, such as Abzakh Adyghe, where *psə* colexifies 'water,' 'juice,' and 'river,' for which latter *psə-x°e* may also be used (*x°e* is glossed as "être, devenir, advenir, augmenter, mûrir" in the source), or Wintu, where *mem* means both 'river' and 'water' and *bohem mem ċuha·* is used for river, *bohe* meaning 'to be big' and *ċuha·* 'to flow.'

Alternatively, there are also complex terms for 'river' where one of the constituents is 'water,' and also here there are cross-linguistically recurrent patterns, and in fact, the optional Wintu complex term already points to two of them: complex terms where the second constituent is 'big' are found in Noni, Santiago Mexquititlan Otomí, Yana (this term also refers to the Sacramento River in particular), and Yaqui, where *batwe* is perhaps analyzable as /ba'a-bwe'u/ 'water-big.' Malagasy *reniràno*, analyzable as /rèny-ràno/ 'mother-water' is very likely an instance of the pattern as well, given the augmentative function of terms for 'mother' in many languages (Matisoff 1992) and the corresponding grammaticalization path; compare also Wappo *ʔéču tùča* /ʔéċu túǎ/ 'creek big.' In some Indo-European languages, terms for river go back to a term for 'water' or more specifically, 'flowing water' (Buck 1949: 41). Central Yup'ik *kuik* is said to contain the "deep root" *ku-* 'flowing liquid,' and terms for 'river' in which the notions of 'flowing' or 'running' figure in addition to 'water' are found in Fijian, Samoan, and Tetun (where 'stream' and 'torrent' are colexified; note also the redundant term *y-syry* 'river/water-flow' in Guaraní). In one sampled language, the word for 'river' and 'flowing water' is of the derived type, namely Cahuilla *wáni-š*, derived from *-wáne-* 'to flow.'

'Way,' 'trail,' or 'track' as the second element in complex terms is found in Laz, Carrier (where 'track' is colexified with 'trace, vestige' and 'site'), and Piro (note also Wayampi *ɨa-la-pɛ* 'canoe-of-way,' and that Lengua *thlinga wathuk* 'stream' contains *thlinga* 'movement,' as well as that Kashaya *biʔda* contains the root *ʔda* meaning 'extend, stretch' which is also found in the term for 'road,' *hiʔda*). This association is realized formally by colexification in Toaripi ("because the rivers and creeks are the highways through much of the low lying Elema countryside;" the language also colexifies 'manner' and 'method,' compare section 92, and, uniquely, 'hand' and 'arm'), and Cashinahua. Moreover, Lesser Antillean Creole French *lawivie* could be analyzed as containing *lawi* 'street,' but it is more

likely to go back to French *la rivière*. Itzaj has *ok ja'* 'leg/foot water/lake/rain' (compare the Yay term *ka*[1] *ta*[5] 'leg river,' denoting a "very large river" specifically), Pawnee *kic-ka* 'be.liquid-on.horizontal.surface' (this term also denotes the Arkansas river specifically; alongside Pawnee, there are also other languages in which the term for 'river' is ambiguous between 'river' in general and a particular river: Dongolese Nubian *úru* also refers to the Nile specifically and assumes the meaning 'to wash out, rinse' as a verb, Badaga *gangi ~ gange* also to the 'Ganges' and the goddess Parvati, Bororo *oroaribo* also to the Rio Paraguay and the Rio São Lourenço in Mato Grosso alongside its capability to refer to a certain spirit, and Mandarin *he2* also to the Huanghe), Maxakalí *kõnãgkox,* analyzable as /kõnã'ãg-kox/ 'water-hole,' Hawaiian *kaha-wai* 'water-place' and *muli-wai* 'after/last-water' (which also means 'estuary'), and Manange *maʃaŋ 2kju* 'low.river.valley water' (the association with 'valley' is also found in Khalkha, Nez Perce, Bora, Huambisa, and Hawaiian by colexification. Note further the possible etymological connection between Basque *ibai*, from earlier **hibaie*, with *ibar* 'valley'). Similarly, Kwoma, Badaga, Itzaj, and Rotuman colexify 'river' with either 'channel,' 'drain,' 'ditch,' or 'gully.' Badaga also colexifies 'bottomland, lowest spot, depression,' and Mandarin also 'plain.'

Also note the similarity between Tasmanian *liapota* 'river' and *liena* 'water,' leading Plomley (1976: 372) to connect the two. Furthermore, Bwe Karen has *chí-bú* 'water-in,' and semianalyzable terms in which one constituent is 'water' exist in Piro, Rama, and Kapingamarangi.

There are also associations exclusively realized by colexification in the languages of the sample: Buli, Hausa (alongside some highly specialized culture-related meanings), Khalkha, San Lucas Quiaviní Zapotec, and Hawaiian (by the analyzable term *kahena wai* 'flowing water') use the same term for 'river' and 'river bed,' and in five languages, Efik, Ngambay, Kashaya, Arabela, and Bora, 'river' or 'big river' and 'ocean' (but with the exception of Efik and Kashaya not 'water'!) are colexified. Waris, Kosarek Yale, Pipil, and Jarawara colexify 'water' and 'river' with 'rain,' while Yir Yoront and Ancash Quechua colexify 'river' with 'Milky Way' (for complex terms for the Milky Way involving a constituent meaning 'river' compare section 37). In Ket, 'water,' 'liquid,' as well as 'alcoholic beverage' are colexified, and in White Hmong, 'water,' 'river,' and 'wine' are.

Other associations include: Efik *uquä* can also refer to a 'flood,' and *akpa'* in the same language also means 'first.' Hausa *kogi* is also used as an epithet as well as the name of a children's game, and Khoekhoe (Haillom dialect) *dommi* also means 'throat, voice.' Ngambay colexifies 'waves.' The Burarra term *angartcha* is derived from the verb *gartcha* 'be stuck' by prefixing of the noun class marker *an-*. Muna *laa* also means 'stem, stalk' and 'straight' inter alia, and another Muna term, *oe*, also is used metaphorically for "interest (in money)." Meyah *mei* also means 'sperm' (alongside 'water'). Rotokas *gae* 'waterway, river' has a verbal reading "follow a course, heed talk, drift, wander." Badaga *oḷe ~ hoḷe* can also mean 'reservoir' and 'swamp,' while Basque *ibai* can also refer metaphorically to an 'enormous lot.' The Khalkha term *ɣulduril* is derived from the verb *ɣulduri-* meaning 'to spill, to be poured out, to pass through' (alongside other meanings colexified) by means of the abstract nominalizer *-il*. Chickasaw *abookoshi* contains *oshi'* 'son,' a morpheme widely used in this language to form metaphorical expression usually conveying a meaning of

smallness. Kiowa colexifies 'river' with 'moon' (the relevant term is also a name for a game). The term for 'river' in Santiago Mexquititlan Otomí, *hñe*, is also the word for 'mirror' (as well as for 'to put on'). Wintu *mem*, colexifying 'river' with 'water,' also means 'wet,' 'thirst,' and 'to baptize.' Aguaruna *namák(a)* also means 'fish.' The Macaguán term *pemnát* also means 'tube,' and Rama *ri* is also used adjectivally with the meaning 'wet.' Hani *lolbaq* contains *lol*, a classifier for rivers, and *baq* means 'direction' or 'thin' and acts itself as a classifier for the side of a mountain and pages of books. Mandarin *he2* also means 'peace, harmony' (with different etyma, Pulleyblank 1991: 122) inter alia, *xi1* 'small river in mountains' also 'to suck in,' 'knee' and 'tin' (all reflecting different etyma, Pulleyblank 1991: 328-330), and *chuan1* also "to bore through, pierce" (both indentical segmentally already in Early Middle Chinese, Pulleyblank 1991: 60). Sedang colexifies 'large river' with 'to imprison,' and Yay with "to put (ones's own money, goods) with another's larger amount" and 'to listen, hear.'

48. The River Bed

Representation: 26%
Motivated: 61.5%
Thereof Analyzable: 40.2% Thereof Colexifying: 25.2%
Thereof by Contiguity: 35.9% Thereof by Similarity: 7.7%
Recurrent associated meanings: river/stream, valley, water, way, flow, base/basis, deep, hole/hollow, place

Lexico-semantic associations for this concept are manifold. Frequently, either 'river' or 'water' is one of the constituents in analyzable terms, though note that five sampled languages, Buli, Hausa, Khalkha, San Lucas Quiaviní Zapotec, and Hawaiian (by the analyzable term *kahena wai* 'flowing water') directly colexify 'river, stream' with 'river bed' (Nez Perce furthermore with 'waterway') and five others, Sentani, Abzakh Adyghe, Khalkha, Lengua, Pawnee (by a term literally translatable as "enclosure on a surface"), and Rotuman, colexify it with 'valley' or 'ravine' (Abzakh Adyghe also colexifies 'precipice,' and Rotuman also 'gutter, gully' and 'channel, trench'), while the Khalkha term has still other meanings and inter alia rarely also assumes the meaning 'large lake.' In Sentani, the relevant term *jaba* 'dry river bed' appears to contain *ja* 'sink,' for which compare Toba *ca'amgue* 'dry riverbed' with *ca'amgui* 'sunk, to lower in the middle.' Otherwise, like for 'river' itself, there are associations with 'way,' 'road,' or 'track' realized by analyzable terms in Mali (*arenggi atha iska* 'river her road'), Nez Perce (*wé·le?skit* /wé·le-ʔískit/ 'flow way;' for this term compare also Wintu *me·m č'oh-i* 'water/river flow-NOMINAL.STEM.FORMANT'), and Miskito (*tingni bila* 'river way' and *li bila* 'water way;' *bila* colexifies many more meanings alongside 'way'). Also note in this context Guaraní *ysyry-ha*, probably analyzable as /ysyry-(a)ha/ 'river go' as well as Nivkh *er myɣ dif* 'river go.downstream track' which is only perhaps so analyzable, but would fit this pattern. Two languages of Eurasia, Ket and Kolyma Yukaghir, have terms for 'river bed' containing an element with the meaning 'deep' (*sést hóbaŋ*, containing *sēs* 'river,' *hòq* 'deep,' and *baʔŋ* 'place,' and *unuŋ-čeginmə ~ unuŋ-čiginmə* 'river depth' respectively), while in two Austronesian languages, Fijian and Tetun, associa-

tions with 'hole' or 'hollow' are encountered (*dreke-ni-wai* 'hollow/cavity-POSS-water' and *mota-kuak* 'river-hole' respectively). In addition, an association with the meaning 'base' is found by colexification in Hawaiian (alongside other colexified meanings, the relevant term *papakū* probably contains *papa* 'flat surface, reef, layer' and *kū* 'to stand;' both constituents also have other meanings) and by the analyzable term *isalẹ odò* 'bottom/base river' in Yoruba. The metaphorical transfer from 'bed' to 'river bed' appears to be peculiar to Europe, occurring in the sample only by the analyzable term *gwely afon* 'bed river' in Welsh. There are also languages in which body-part metaphors are employed to convey the meaning 'river bed': 'guts' and 'river bed' (as well as 'inside') are colexified in Jarawara, Khoekhoe has the analyzable term *!ā-!nā-b* 'river-stomach/interior-3SG.MASC,' and Rendille colexifies 'arm, hand,' 'elephant's trunk,' and 'river bed.' Tetun *mota-fatin* is analyzable as 'river-place,' compare the semianalyzable Carrier term *nethayiḳĕt* containing *ḳĕt* 'place,' that the second element in the Cavineña term *ejiri quini* means 'broad place,' and that Pawnee *huukahaaru*ʔ, denoting a 'dry riverbed' specifically and at the same time colexifying 'beach, shore' and "dry bed in a pond," is literally "in water place." Pawnee also has the term *huukaahaaru*ʔ, containing *huuka(wi)-* 'along a stream course' and *haar* 'place;' this term also means 'beach, shore.' Other morphologically complex terms with 'river' include Efik *isöñ akpa* 'earth/ground river,' Rotokas *uuko gae*, containing *gae* 'river' and *uuko*, 'liquid state, fluid,' 'to collect water,'' Sora *əluŋˈnai* /ə-luˑŋ-ˈnad/ 'POSS-sleep.or.soak.in.water-river,' Chickasaw *abookoshi' shila-'* 'river dry-NMLZ,' denoting a dry river bed specifically. Hawaiian colexifies 'riverbed' with 'bottom' among other meanings. Furthermore, Samoan *'alitivai* contains *vai* 'water' but is not amenable to full morphological analysis, the same situation is encountered in Piro. There is a semianalyzable term featuring a constituent 'river' in Yuki. Finally, Sedang *chúa* is also the name of a weaving design inter alia.

49. The Root

Representation: 97%
Motivated: 32.6%
Thereof Analyzable: 6.9% Thereof Colexifying: 25.8%
Thereof by Contiguity: 5.4% Thereof by Similarity: 8.4%
Recurrent associated meanings: base/basis, tendon, origin, trunk, reason/cause, vein, nerve, muscle, stump, tuber, foot, leg, branch, buttocks, liana, root of tooth, tongue root, stalk, foot of hill, tendril, thread, tree

The 'root' is a meaning expressed in many languages by non-motivated, non-analyzable terms. If they are motivated, however, body-part metaphors are used in some languages to express the concept. In three sampled languages, Buli, Burarra, and Miskito, 'root' is colexified with 'foot,' and in Abzakh Adyghe, *λapse* is analyzable as /λ(e)-ps(e)/ 'foot-string.' In Buli, Ngaanyatjarra and Yir Yoront, 'root' is colexified with 'leg' (in Buli, by the same term that also colexifies 'foot;' it also means 'branch' inter alia, as does the relevant Ngaanyatjarra term which denotes a 'side root' specifically. Buck 1949: 522 also reports this association for Indo-European. Yir Yoront also colexifies 'tail' of certain fish species

and is used generically for certain molluscs). Bwe Karen has *θo-kha-wi* 'tree-leg-vein.' There are also some languages in the sample in which 'root' and 'vein' are colexified, namely Rendille, One, Basque, Carib, Lengua, and Miskito (similarly, Gurindji more specifically colexifies 'single root of tree' with 'vein,' and Hawaiian 'small root, rootlet'). However, for this association, cross-linguistic tendencies in the directionality of the mapping are hard to assess, given that in Kanuri, 'vein' is *zâr bû-bè* 'root blood-of.' Many sampled languages also colexify 'root' with 'tendon.' This association is found in Kwoma (where also 'tendril' is colexified, a pattern shared with Kiliwa), One, Chickasaw, Ineseño Chumash (also meaning, presumably by provenience contiguity, 'bowstring'), Pawnee, Jarawara, Lengua, Miskito, and Hawaiian. In addition, 'root' and 'muscle' are colexified in Koyraboro Senni, Ineseño Chumash, Jarawara, and Hawaiian (where the relevant term conflating all three meanings is also used figuratively for 'womb' and 'offspring'), and 'nerve' in Koyraboro Senni, Basque, Ineseño Chumash, Miskito, and Hawaiian. Note that there is a large overlap between the four groups due to the fact that very frequently 'tendon' and 'vein,' and somewhat less frequently also 'muscle' and 'nerve' are colexified (see sections 141 and 147). The Koyraboro Senni and Highland Chontal terms also denote a 'liana, vine' or species of liana. Two languages, Sora and Rotuman, colexify 'root' with 'root of tooth,' Hausa and Yoruba colexify it with 'buttocks,' and another two, Cashinahua and Mandarin, with 'tongue root.' Sko *hangling* might be based on a metaphorical comparison with *háng* 'end of intestines, kidney' (the second element of the term is unknown, and a similar situation is encountered in Kwoma; for this possible association, compare Great Andamanese *ârchâg* 'root' with *ôngchâg* 'kidney,' presumably motivated by alternation of noun class for which compare discussion in § 4.1.1.2.). Lavukaleve *kala* also means 'collarbone,' and for Sentani *kambu*, compare *kambi* 'neck.' The Wintu term *ċaraw ẋosi* contains *ċaraw* 'green land, field, valley,' and *ẋosi* might be related to *ẋOs* 'fog, steam, gas, lungs.' Kaingang *jã-re* appears to be analyzable as 'tooth-field,' and Lenakel *nuk-* may also refer to the 'armpit.'

Intra-domain ties are somewhat less frequent. Seven languages, Buli, Abzakh Adyghe (by the analyzable term mentioned above, note that this term also means 'thread' and compare Swahili *mzizi*, analyzable as /m-uzi/ '3/4-thread'), Yaqui, Lesser Antillean Creole French, Abipón, Miskito, and Tehuelche colexify 'root' with 'trunk' or parts thereof (Miskito also with 'pole' and 'substance,' and Tehuelche also with 'handle of knife,' 'claw,' and 'behind of'). Yaqui does so also with 'stump,' an association it shares with Biloxi, Lesser Antillean Creole French (where the term also means 'stem') and Nez Perce. The Hausa, Badaga, Pipil (Santo Domingo de Guzmán dialect), and Jarawara terms are also used with the meaning 'tuber.' Badaga *ga:su* is also used specifically with reference to the 'potato,' and Tuscarora *uhné?reh* is extended to 'turnip' and 'vegetable' in general. Another association is that between 'root' and 'stalk,' found in Abzakh Adyghe and Badaga (in the latter language, by a semianalyzable term containing a root meaning 'to grow,' 'fertile,' and 'yield' inter alia, and itself colexifying "water dripping from roots").

There are also a number of associations where the terms for 'root' are used to map some more abstract meaning. Buli, Hausa, Kyaka, Abzakh Adyghe, Khalkha, Kolyma Yukaghir, Ancash Quechua, Fijian, and Malagasy colexify 'root' with 'base, basis' (Kolyma Yukaghir also with 'custom' and 'similarity'), 'origin' is a secondary meaning of 'root'-

terms in Buli (where it is restricted), Anggor (where the term *ahasahari* is semianalyzable containing *ahari* 'stem' and colexifies 'clan'), Basque (where one of the relevant terms also means 'root in mathematics,' 'stock, lineage,' 'tentacle,' 'udder,' 'hinge,' 'inclination, tendency,' 'beam, sunbeam' and 'segment'), Khalkha (also expressing the notions of principalilty and originality, among others), Nez Perce, Lesser Antillean Creole French, Embera, Rotuman, and Tetun. 'Reason' or 'cause' is colexified in Buli, Nez Perce, Hawaiian, Malagasy, Mandarin, and Rotuman (the relevant Buli term participating in these patterns is the same that also colexifies 'trunk,' but a different one from that realizing the associations with 'leg' and 'root' reported above, the Hawaiian term also means 'foundation' inter alia, and the Rotuman term is also used with the meanings 'principal village, capital,' 'lower end,' and others). Another metaphorical association, that with 'foot of hill/mountain,' is attested in Abzakh Adyghe and Greek.

Further less systematic associations include: Buli *kiri* can also refer to the 'character' of a person, 'type,' as well as 'under, below' (compare the colexification of 'root' with 'bottom' in Hawaiian, and that with "upward from below" in Kapingamarangi). Efik *äduñ* appears to be derived from *duñ* 'to dwell, inhabit.' Hausa *saiwa* can also refer to a "drabbish-coloured goat," and another Hausa term, *tushe*, colexifies 'root of plant' with 'plant itself' inter alia. For Khoekhoe, it is noted that *!nomab* also means 'root' in the linguistic sense. Oddly, Ngambay *ndìrà-kake* seems analyzable as 'hard.to.eat-tree,' while *ko*, another term for 'root' in the same language, also means 'to cry' and 'seed,' and *ngìrà* can also be used adjectivally with the meaning 'hard.' Noni *gbweŋ* also denotes the 'stinger of a bee,' and Rendille *híy* also means 'sour milk' and 'relatives, kin.' Baruya colexifies "fine root in or out of ground" with "fat from a pig's stomach," while Berik *tiskar* and Dadibi *ni pedali* 'tree root' apparently contain *ti* and respectively *ni* 'tree, wood.' Kwoma colexifies 'root' inter alia with 'wrinkled,' and Gurindji *wirnturru* also means 'horn.' The Muna term *paraka* is also the name of an evil spirit that eats people. Meyah *ofóm* also means 'ripe.' The Rotokas term *vavu-rupa-to* is analyzable as 'bitter-dark-SG.M,' while Sahu *utu'u* also means 'buttress' specifically. A Badaga term for 'root,' *be:ru*, contains the verb *be:* 'to grow' and colexifies 'finger.' Khalkha *yndysy(n)* also means 'race, nationality' and is the name for the religious writings in the 'Tantra,' and Welsh *gwreiddyn* is also used with the meaning 'stock.' Chickasaw *ishtaahikki'ya'* is analyzable as /isht aa-hikki'ya-'/ 'with LOC-stand-NMLZ' (the term refers to above-ground roots specifically), Itzaj colexifies "hateful, angry," Wintu 'herb charm' by a semianalyzbale term containing a constituent referring to level land and *ċe·k* colexifies 'ropelike root(s)' with 'rope, cord string,' 'to tie a rope.' The Yuki term *koot´kin ~ kutkin* might be derived from the verb *koot´* 'to start.' A Central Yup'ik word for 'root,' from the Norton Sound dialect, *acilquq*, is derived from the noun *aci* 'area below, area under' by suffixation of the postbase (see § 4.4.2) *-quq*, meaning 'one that is like' (there is the variant *acipluk* /aci-lluk/ 'area.below/area.under-bad'). Another Central Yup'ik term, *nemernaq*, consists of the noun *nemeq* 'binding, wrapping' and the postbase *-naq* 'one like;' this term also denotes the 'lamprey.' Cashinahua *tapun* might contain *tapu* 'platform, shelf, table,' and Cayapa *telele* 'root' *te* 'firewood.' Chayahuita *itë'* appears to be analyzable as /i'-të'/ 'water-INSTR,' while Macaguán *petakomét* appears to be analyzable as /peták-omét/ 'container-sun/moon.' Piro colexifies 'pelvic bones,' and Maxakalí

mĩmyĩpxatit contains *mĩm* 'wood.' Toba *lpa'a' ~ lpa'a'q* is apparently related to *pa'a'*, which is the name of a plant the root of which causes hallucinations. Fijian *vū* also means 'cough, to cough' and 'to wash, cleanse,' and Hani *alqil* and *daoqqil* might be related to *qil* 'be firm, be durable.' Figuratively, Hawaiian *a'a* also means 'womb, offspring,' as well as "to send greetings of love; joyous hospitality; joy at greeting a loved one." Finally, Samoan *a'a* also means 'connection, "involve, implicate," and 'influence,' Sedang *rei* also is the name of a bird inter alia, while Yay *raak*[5] also means 'to pull, drag.'

50. The Seed

Representation: 89%
Motivated: 43.3%
Thereof Analyzable: 6.4% Thereof Colexifying: 36.9%
Thereof by Contiguity: 5.7% Thereof by Similarity: 13.1%
Recurrent associated meanings: fruit, egg, offspring/descendants, eye, testicle, semen, plant/sow, round/round thing, nut, bead, germ, heart, stone, breed, bullet, bone, plant, grow, kind, bud, face, product, peer/friend, tuber, berry, flower

Associations between 'seed' (or 'grain' / 'pit') and other small roundish objects abound cross-linguistically. Languages of the New Guinea area (but also some others) are particularly fond of colexifying 'seed' with other products of plants or animals that have the property of being round. Buin, Kyaka, Rotokas, Kosarek Yale, Abzakh Adyghe (inter alia), Wintu, Hawaiian, and Takia colexify 'seed' with 'egg,' and Gurindji, Kwoma, Kyaka, Ket (colexifying also 'pine nuts' specfically), Kapingamarangi, and Takia colexify 'seed' with 'nut.' More generally, 'seed' is associated with 'fruit' in many languages as well, namely in Efik (here also with 'flower' by a term derived from a verb meaning 'to peel' inter alia; in Waris, 'flower of trees' is specifically colexified), Ngambay, Dongolese Nubian, Buin, Kwoma (here also with "edible part of a plant"), Kosarek Yale, Khalkha, Kiowa (colexifying also 'vegetable' and 'bread'), Abipón, Miskito, Piro (where the relevant term *xi* also acts as a diminutive marker), Rama (colexifying 'peanut' specifically), Sáliba (also colexifying 'bud,' as is the case in Efik and Lesser Antillean Creole French), Wayampi, Hawaiian, Takia, Sedang (colexifying also 'pellet in blow gun' and 'muscle'), and Yay. Similarly, Tetun colexifies 'berry,' while Cheyenne *hestáhame* is literally 'heart berry.' Takia, alongside all these meanings, also colexifies 'shell.' This association may be based on meronymy in some cases, but perceptual similarity might also be at work, since many fruits are globular as well. In Rama, the term colexifiying 'seed' with' fruit' is *kat up* 'tree eye,' and this illustrates that associations with other round objects of small size also transcend domain boundaries. Pawnee is another language with a complex term for 'seed' (in particular 'planting seed') on the basis of 'eye' (*rak-kirik-u*ʔ 'wood-eye-NOM'), and the association is realized by colexification in Anggor, Burarra, Cahuilla (also colexifying 'face'), Chayahuita (also colexifying 'design of a waistband'), Jarawara (here also colexifying 'face,' 'pile,' 'color,' 'end,' and 'pellet'), and Lengua, while the Nunivak Island dialect of Central Yup'ik and Huambisa have seminanalyzable terms. For Burarra *mipila*, the dictionary gloss reads as follows: "eye, or anything suggestive of an eye by virtue of being small and round or exud-

ing fluid, hence hail, seed, bullet, hook of woomera, glass tumbler, nipple, spring of water," suggesting that 'eye' is the dominant reading (see also § 6.2.3.1.). The extension to 'bullet' is also found in Ngambay, Toba, and Yanomámi (here also 'almond' is colexified'); note that 'gunshot' is colexified in Lengua. Four languages, Muna, Badaga, Welsh, and Maxakalí have terms colexifying 'seed' with '(small round) stone,' and four languages in the sample, Sentani, Kosarek Yale, Nez Perce, and Samoan colexify 'seed' with 'heart' (Nez Perce also with 'pith'). In addition, Cheyenne *hestáhame* literally means 'heart berry,' which suggests a metaphorical transfer of the position of the heart in the middle of the chest to the position of the seed in the center of a fruit, corroborated by the frequent association between 'heart' and 'middle' (see section 117, though note that the transfer seems to go in the other direction in Hupda *hã́wɨg* which probably goes back to **haŋ-wɨg* 'breath-seed'). Transfer from 'seed' to the domain of body-parts also occurs with another meaning: seven sampled languages, Welsh, San Mateo del Mar Huave, San Lucas Quiaviní Zapotec, Guaraní, Toba, Hawaiian and Kapingamarangi, colexify 'seed' with 'testicle' (Yanomámi with 'penis'), and about the same number of languages have complex terms for 'testicle' based on 'seed' (compare section 142 as well as Brown and Witkowski 1981 for further evidence). Colexification with 'bead' is found in six sampled languages, Basque, Lesser Antillean Creole French, Central Yup'ik (which also colexifies 'single fish egg' and "any other seed-like thing"), Imbabura Quechua, Maxakalí, and Vietnamese. Moreover, there are also some sampled languages where a term for 'seed' in fact is glossed as having a secondary meaning of 'round thing' directly. These are Berik (the term seems to be semianalyzable, containing an element meaning 'tree, wood'), Muna (where the relevant term curiously acts as a classifier also for "any big object;" this term means also 'to be present, gather together'), One, Hupda, and Hawaiian (note also that Badaga *guṇḍu* assumes the meaning 'round' and 'stout, strong' when used as an adjective). Less frequent patterns involving colexification of 'seed' with a roundish object include that with 'marble' in Buli, with 'nipple' in Gurindji, with 'tuber' (as well as 'electric battery' and other meanings) in Kyaka and Hawaiian, with 'flower in bud' in Sáliba, with 'bulb' in Hawaiian, with 'clitoris' in Muna (cf. the extension of 'fruit' to 'clitoris' in Australian languages, Austin et al. 1976), with 'tablet' in Ngaanyatjarra, and with 'moon, month' inter alia in Rotuman.

Other recurrent associations are that with 'bone' in Yir Yoront, Basque (here, also with 'fishbone'), Khalkha, and Wintu (common features: hardness and the fact that both 'seed' and 'bone' are found within larger structures, the body and the fruit respectively? at least the former is suggested by the fact that 'hard, solid' alongside 'unyielding' is also colexified in Wintu), that with 'kind' in three languages of different families of Africa, Efik (here only as of animals and vegetables), Hausa, and Koyraboro Senni (note that Hawaiian *'ano'ano* might be reduplicated from *'ano*, one of the meanings of which is 'kind, type'), that with 'offspring, descendants' (sometimes also 'child') in Hausa, Khoekhoe, Dongolese Nubian, Kosarek Yale, Khalkha, Wintu, Hawaiian (in both Khalkha and Hawaiian also with 'product;' see Buck 1949: 505 for the same extension in Swedish), and Bwe Karen, which colexifies also 'nine.' Similarly, Badaga *bittu* colexifies 'father's line, patrilineage,' as well as 'yield' and 'crop' with 'seed,' and the Guaraní term colexifying 'semen' and 'testicles,'

ta'ŷi, is analyzable as /ta'y-i/ 'son/clot-DIM.' Still other recurrent patterns of colexification include that with 'breed' (e.g of animals) in Efik, Basque, Miskito, and Tsafiki (similarly, Khalkha inter alia colexifies 'race, family, clan'), that with 'peer, friend' in Koyraboro Senni and Kyaka, and that with 'semen' in Dongolese Nubian, Basque, Greek, Welsh, Wintu, Aymara, and Guaraní (note that both referents are similar in that they are part of the reproductive system, and see also Buck 1949: 505 for Indo-European, in particular Swedish). Two of these languages, Greek and Guaraní, as well as Kyaka, Lesser Antillean Creole French and Hawaiian, also colexify 'germ;' in the context of the association with 'semen,' note also Kyaka *wai*, glossed as 'germs, spark of life' alongside 'seed' (as an adjective, *wai* means 'introduced, not native, not local'). Ineseño Chumash *'amɨ'n* also means 'body' and 'flesh, meat' alongside 'seed,' and Hawaiian colexifies "meat as in 'opihi shell or 'alamihi crab."

Entirely different structures are found in terms for 'seed' when they are primarily conceived of from the point of view of agriculture. Then, terms derived from or containg verbs meaning 'to plant' or 'to sow' are common (as is the case diachronically in Indo-European, Buck 1949: 505). Koyraboro Senni and Badaga colexify these meanings, and fully analyzable terms with such structure are found in Ngambay, Abzakh Adyghe (colexifying 'cereal' generally), Cheyenne, Nez Perce, Xicotepec de Juárez Totonac, and Aguaruna, which latter for instance has *ajákma-mu* /ajákmat-mu/ 'sow-NMLZ.' Semianalyzable terms suggesting such a structure are also attested in Koyraboro Senni and Dongolese Nubian, and the association is recoverable diachronically in Greek. In connection with this pattern, note also Carrier *hananelyih-î* 'grow.again-REL' and Nivkh *vandu oxt* 'grow powder,' and colexification of 'seed' and 'to grow' in Basque and Abzakh Adyghe inter alia. A similar account is available for Quileute *ḳatsàḳółwa,* containing *ḳátsa-* 'to bury;' moreover, Baruya, Aguaruna and Huambisa colexify 'plant' and 'seed,' and Upper Chehalis colexifies 'seed' with 'plant' as well as with 'garden.'

Other associations are: Buli *biri* also means 'counter' as of a particular game, Hausa *iri* also 'slips,' and Ngambay *kànde* also 'to produce' as well as 'genetic inheritance' inter alia. *Ko*, another term from the same language also means 'to cry' and 'root.' Koyraboro Senni *dumari* also means 'sprout,' and in Rendille 'seed' and 'tear, teardrop' are colexified. Baruya *wia* also means 'seedling,' and Kwoma *siik,* alongside 'seed,' 'nut,' and 'fruit,' may also refer to a 'clot of blood' inter alia. Kyaka *kapa* not only participates in the colexification patterns with 'nut,' 'egg,' and 'peer,' but also may refer to the 'core' or 'nucleus' of something in general, as well as to a 'larva,' and 'fat, suet, grease' inter alia. Kosarek Yale *wana* can also refer to a 'flower-stalk.' The Muna term *lumu* also means "seedpod inside fruits" and 'moss,' and One *tala* also 'tree stump' and 'grasshopper,' while Abzakh Adyghe colexifies 'seed' with 'fur, feather' and 'tooth,' and the Khalkha term *køryngge(n)* also is used inter alia to refer to 'yeast' and 'property, resources.' Biloxi *su* also means 'blown out, extinguished,' while Upper Chehalis *smámɨs* is derived from the verb root *náma-* 'done, finished, quit.' Kiowa *dǫųgā't* is literally translated "inside one," Oneida colexifies 'seed' with 'oats,' while Santiago Mexquititlan Otomí *nda* also denotes 'cotton' specifically. Pawnee *riikactikiisuˀ* is analyzable as /riikac-rikiis-uˀ/ 'crookneck.squash-kernel-NOM' (the term is also used generically for 'seeds' of all kinds), and Wintu colexifies

'seeds' with 'teeth' by one term and with 'stab, pierce, poke' by another. Carib *epɨpo* is derived from *epɨ(lɨ)* 'flower' and means also 'stem' and 'stick,' Cayapa *ñi* also means 'flame, fire' (probably due to accidental homonymy), Kaingang colexifies 'seed' with 'braid, bolt,' Miskito with "bulk, mass, lump, particle" inter alia, Ancash Quechua with 'black and white mottled,' Tehuelche with 'leaf,' and Wayampi with 'foot' inter alia and by another term with 'almond.' Finally, Yanomámi *mo* colexifies 'seed' and 'penis.'

51. The Shadow

Representation: 92%
Motivated: 49.3%
Thereof Analyzable: 7.7% Thereof Colexifying: 42.4%
Thereof by Contiguity: 7.2% Thereof by Similarity: 37.8%
Recurrent associated meanings: soul/spirit/ghost, reflection/image/mirror, image/picture/drawing, photograph, shelter, dark/darkness, cold/cool, cloud

Very frequently, 'shadow' (or 'shade') is associated with 'soul,' 'spirit' or 'ghost' lexically (or, in this case, more likely primarily culturally or mythologically). This association is particularly common in Oceania and the Americas (though reported by Buck 1949: 62 to be common in Indo-European "from Homer on"), occurring by colexification in Mbum, Efik, Ngambay, Anggor (where the relevant term also denotes a type of fly), Burarra, Kaluli, Kwoma, Kyaka, Lavukaleve, Rotokas, Toaripi, Sentani, Sko, Kosarek Yale, Greek, Kolyma Yukaghir, Blackfoot, Cheyenne, Pawnee, Quileute, Wintu (colexifying also 'glimpse' inter alia), Abipón, Aguaruna, Arabela, Carib, Cayapa, Chayahuita, Jarawara, Kaingang, Maxakalí, Miskito, Ancash Quechua, Rama, Wayampi, and Lenakel (35 languages), and by complex terms in Meyah (*efená órka* 'spirit carry'), the Nunivak Island dialect of Central Yup'ik (*tarenraq* /tarneq-aq/ 'soul/spirit-thing.that.resembles.in.some.respect'), and Fijian (*yaloyalo-na* 'reflection-POSS,' with *yaloyalo* reduplicated from *yalo* 'soul, spirit'). Tsafiki, in addition, has a semianalyzable term containing *o'có* 'evil spirit' and colexifying 'firefly.'

There are a number of further associations which cluster strikingly in Oceania and more specifically in New Guinea. Colexification with 'mirror,' '(mirror) image,' or 'reflection' is attested in Hausa (*inuwa*, the relevant term, is also a "name given to any one called Muhammadu"), Baruya, Burarra, Kaluli, Kwoma, Kyaka, Lavukaleve, Mali, Muna (where the relevant term is also the name of a tree species), Nunggubuyu, Rotokas, Sentani, Toaripi, Kosarek Yale, Itzaj (colexifying also 'nature'), Lake Miwok, Central Yup'ik (Nunivak Island dialect), Arabela, Bora, Bororo, Cayapa, Kaingang, Piro, Rama, Yanomámi, Bislama, Hawaiian (also colexifying 'embryo,' 'newly hatched fish,' knuckles' and 'joint,' inter alia by one of the relevant terms, and 'bright,' 'dazzling,' 'white' and similar meanings by another), Lenakel, Mandarin, Rotuman, Samoan, Sedang, and Yay, in which latter the relevant term also means 'to shine, sheen' (33 languages); moreover, in Fijian, the association is by the analyzable term mentioned above. Colexification with 'picture, image' and/or 'drawing' is attested in Bakueri, Efik (by the term *ñwet ~ ñwed* derived from *wet* meaning 'paint, write, mark' inter alia, the derived term also meaning 'pattern,' 'inscription' as well as 'writing' and, presumably by further semantic extension, 'book' inter alia),

Ngambay, Noni, Rendille, Burarra, Kwoma, Mali, Nunggubuyu, One, Rotokas, Toaripi, Sahu, Sko (by the term, *bàleng*, presumably analyzable as /bà-lèng/ 'person-hide.self'), Kosarek Yale, Yir Yoront, Abzakh Adyghe, Central Yup'ik (Nunivak Island dialect), Abipón, Bora, Maxakalí, Miskito (also colexifying 'resemblance'), Sáliba (also colexifying 'appearance' and 'color'), Fijian, Hawaiian, Lenakel, Rotuman, and Samoan (28 languages), and with 'photograph' in Ngambay, Noni, Rendille, Kwoma, Mali, Rotokas, Sko, Khalkha (colexifying also 'age' and 'apoplexy'), Central Yup'ik, Arabela, Bora, Jarawara, Lenakel, Rotuman, and Samoan (15 languages, note that with the exception of Khalkha and Jarawara, languages which betray this association actually are a subset of languages which colexify 'picture').

In contrast, six sampled languages reveal a different conception of 'shadow' in that the relevant terms bear an association with 'dark' or 'darkness' (as also evidenced by an etymological connection between German and Greek in Indo-European, see Buck 1949: 63). This is found in Mali, Ngaanyatjarra, Kashaya (which also colexifies "sickness caused by fear") by colexification, while realized by complex terms in Ket (*qon-sal* 'dark-night' and *qonij baˀŋ*, containing *qon* 'dark' and *baˀŋ* 'place') and the Santo Domingo de Guzmán dialect of Pipil (*ku:-yuwa* 'tree-dark'); moreover, a not entirely analyzable term containing *kʻǫų* 'be dark' is found in Kiowa. There is also colexification of 'shadow' with '(to) shelter,' as encountered by colexification in Buli, Ngaanyatjarra (where the relevant term also denotes a 'metal canopy' as well as rings around the moon which are said to indicate coming rain), Welsh, Kapingamarangi (where the relevant term also means 'behind some protective cover') and Rotuman, as well as by the analyzable term *moˈã-ha* 'cover/protect-AGT' in Guaraní (see Buck 1949: 62 for diachronic evidence from Irish; the Guaraní term colexifies also 'defense' and 'protection'). What the structure of these terms show is that here terms are unlikely, as opposed to the examples above, to refer to the shadow of a person specifically. In fact, for instance Efik distinguishes the two lexically: *ukpöñ* is used for the shadow of things that move, and *mfut* for the shadow of things that do not. A similar distinction is made in San Mateo del Mar Huave, although terms for both variants share the same root, with that for persons or animals being inalienably possessed. Baruya has unrelated lexical items for the shadow of clouds and the shadow cast by other entities.

Two other languages, Buli and Arabela, have associations between 'shadow' and 'cold' or 'cool place' (note also Embera *kũrásare* 'shadow' and *kũrása* 'cold;' this association is recoverable etymologically in Donolgese Nubian, see also Buck 1949: 63 for evidence from Baltic and Serbo-Croatian). Moreover, Bororo colexifies 'shadow' with 'cloud,' and Biloxi *si natci* appears to be analyzable as /si natci´/ 'feet cloud' (compare also Chickasaw *hoshontikachi'* 'shadow' with *hoshonti* 'cloud,' and see Plomley 1976: 383 for the possibility of this connection in Tasmanian).

Other unsystematic associations include: the Buli term *yogsum*, colexifying 'shelter' and 'coolness,' also means "danger, fear, fright, dread, apprehension, terror." Efik *mfut'* is apparently derived from *fut*, meaning 'to swell, boil, foam' inter alia, Katcha *bogo* also means 'place,' and Koyraboro Senni *bii* (related diachronically to *bibi* 'black') is also used to refer to an 'umbrella' and also means 'yesterday.' Kaluli colexifies 'shadow' with "reverberation in forest or in memory," Kwoma *mayi* also means 'map' and is furthermore used to refer to a variety of supernatural powers, One *iloula ~ ilwola* also denotes a "meas-

uring stick,' Sahu *gu'dumini* also means 'life strength,' and Western Tasmanian perhaps colexifies 'shadow' with 'fog.' Badaga colexifies 'shadow' with 'shape,' and Basque *itzal* also means 'prison' and 'respect, prestige' inter alia. Kolyma Yukaghir *numet* also means 'fontanel.' Santiago Mexquititlan Otomí *xudi* also means 'morning,' while Oneida colexifies 'shadow' with 'movie, show' (similarly, Fijian colexifies 'film'). Tuscarora *utiʔθrę́hsteh* is analyzable as /u-tiʔθ(e)r-ę-(a)hst-eh/ 'NOUN.PREFIX-overhang-fall-NMLZ-NOUN.SUFFIX,' Bororo *arodu* also means 'bad, false,' 'robbery,' 'sign of disease and death,' as well as 'hole,' Cayapa *aama* also means 'weapon' (presumably due to collapse of Span. *alma* 'soul' and *arma* 'weapon'), and Cashinahua *baka* also means 'scar' as well as 'spouse' and 'friend.' Chayahuita *sanohuan* is derived from *sano* 'quiet' by means of a classifier suffix meaning 'one who has,' while Guaraní *kuarahy'ã* contains *kuarahy* 'sun.' Ancash Quechua *qitqi ~ qetqi* 'shadow of cloud' also means 'soot,' Yanomámi *noreshi* is also used to refer to the "phosphorescent shining of certain plants when decomposing" as well as a "double of humans incarnated in animals." Great Andamanese *ôtlêre* is apparently derived from *lêre* 'black beeswax.' Samoan *ata* also means "copy, duplicate," and, in the plural, "[l]ight and shade effect," while Tetun *mahon* also means "influence, sway" or 'framework.' Yay *ram*[5] is also the name of a 'big hawk that soars in the sky and eats chicken,' Vietnamese *bóng* also means 'ball,' 'cotton,' and 'flower,' wheras Bislama *sado* (< Engl. *shadow*) also means "humiliation, dishonour."

52. The Sky

Representation: 97%
Motivated: 49.1%
Thereof Analyzable: 11.8% Thereof Colexifying: 37.9%
Thereof by Contiguity: 17.1% Thereof by Similarity: 22.7%
Recurrent associated meanings: heaven, high/above/up, cloud, blue, top, air, god, weather/climate, day, light, world, rock, sun, roof, rain, ceiling, hole/opening

The most frequent lexico-semantic association for 'sky' (or 'firmament') is that with 'heaven(s)' by colexification, occurring in 44 languages, namely Buli, Hausa, Khoekhoe, Ngambay, Swahili, Yoruba, Buin (colexifying also 'swollen' inter alia), Dadibi, Kaluli, Kyaka, Ngaanyatjarra, Rotokas, Sentani, Toaripi, Kosarek Yale, Yir Yoront, Badaga, Basque, Chukchi, Ket, Khalkha, Kildin Saami, Highland Chontal, Kiliwa, Lesser Antillean Creole French (which also colexifies 'paradise'), Lake Miwok, Lakhota, Pawnee, Pipil (Santo Domingo de Guzmán dialect), Quileute, Central Yup'ik, San Lucas Quiaviní Zapotec, Carib, Maxakalí, Piro, Hani, Hawaiian, Lenakel, Malagasy, Rotuman, Samoan, Tetun, Yay, and Vietnamese (similarly, Muna colexifies 'sky' with "space between heaven and earth").

The second most frequent association, that with the meanings 'high,' 'above,' or 'up, upward,' has a very striking pan-American distribution in the sample, while being rare elsewhere in the world. This is unexpected, given the seemingly related grammaticalization path 'sky' > 'up' reported by Heine and Kuteva (2002: 279) also attested in languages of other regions of the world. In the sample, the association is realized by colexifi-

cation in Buli, Mbum, Rendille, Upper Chehalis, Ineseño Chumash, Nuuchahnulth, Jarawara, Lengua, Macaguán, and Miskito, but occurs also frequently by morphologically complex terms: Kashaya has *qali q^{h}aʔbe* as 'clear/above rock,' Lake Miwok *líile-wali* 'high-world,' Santiago Mexquititlan Otomí *ma-hets'i* 'LOC-high,' Pipil (Santo Domingo de Guzmán dialect) *ka-ihakhu* 'in/at-high,' Aymara *alajj-pacha* 'above-whole,' Piro *ten-hohne* 'high/tall/deep-expansive,' and Imbabura Quechua *jawa pacha* 'above space.' Moreover, Comanche has the formally redundant term *tomobaʔatɨ̱* /tomo-paʔatɨ̱/ 'cloud/sky-above,' and a semianalyzable terms suggesting such a structure is encountered in Ineseño Chumash.

Several of the associations by terms of the lexical type reported above also occur in other configurations elsewhere. Associations with 'cloud' (which are also frequent in Indo-European, Buck 1949: 52-53) occur by colexification in Yoruba, Lavukaleve, possibly in Nunggubuyu, Bezhta, Upper Chehalis, Cheyenne (with doubts on behalf of lexicographers), Comanche, Kiowa, Lakhota, Nez Perce, Miskito, and Rama (by the analyzable term *núnik kás* 'sun/day meat;' some of the terms have secondary associations due to the 'cloud'-reading, see section 12). Swahili *mbingu* 'sky' consists of *bingu* 'cloud' prefixed with a noun class marker, and Biloxi *natcí tohí* is analyzable as 'cloud blue' (and denotes the 'clear sky' specifically). Moreover, Efik has the analyzable term *ikpa'enyöñ* /ik'pa-en'yöñ/ 'animal.skin-sky/heaven' used for both 'firmament' and 'cloud.' Hawaiian *aouli* "Firmament, sky, blue vault of heaven" contains *ao* meaning inter alia 'light, day, dawn, cloud' (compare the association with 'light' in Swahili noted above, 'light' and 'sky' are also colexified in Guaraní), and *uli* which can refer to dark colours, including dark green and blue. A color term for 'blue,' as in Biloxi, also figures in other languages of North America: Upper Chehalis *tit ʔacq̓^{w}éʻx̣m̓ɬ* is derived from *q̓^{w}íx̣-* 'blue,' and Cheyenne has *otá'tavó-'omëë'e* 'blue-realm;' moreover, Oneida and Tuscarora colexify 'sky' and 'blue.' The association between 'sky' and 'world' in Lake Miwok corresponds to the colexification of these two meanings in Hausa (where it is archaic) and Arabela, which has a term with very wide semantic range, also including 'earth' and 'hole, opening,' which latter meaning is also colexified in Bora. And the Kashaya association with 'rock' is paralleled in Miskito and Hupda, which colexify the meanings (Hupda also colexifies 'mountain'). They mirror precisely the pre-history of the Indo-European inherited term evidenced mainly by Avestan and Sanskrit evidence (Buck 1949: 52).

Four sampled languages, Yoruba, Kosarek Yale, Welsh, Hawaiian (the term here also means 'to float,' 'homeless (person)' inter alia, and denotes a particular star), and Rotuman have terms colexifying 'sky' with 'air.' In four sampled languages of Africa and Eurasia (and in some other Indo-European languages, Buck 1949: 52-53), Bakueri, Buli, Ket, and Khalkha, the word for 'sky' is also the name of (a) god (Bakueri also has *mmányú ya ló̱wa* 'up of god/sky'). 'Sky' is colexified with 'top' in general in five sampled languages, namely Efik (also colexifying 'lift'), Hausa, Koyraboro Senni, Upper Chehalis, and Lengua, and, relatedly, in two languages, namely Hausa and Basque, 'sky' is colexified with 'roof' (the Basque term also may assume the meaning 'canopy,' 'glory,' and 'ceiling;' the latter meaning is also colexified in Central Yup'ik, compare also Bwe Karen *mɔ́kho* 'sky' and *kho* 'top, roof'). In Buli, Khalkha, Wintu, and Rotuman, the relevant term may also assume the

meaning 'weather' and/or 'climate' (Buli also colexifies 'season, period, time' and is a "religious concept denoting the 'alter ego' or 'personal god' of an individual," while in Wintu, 'clear weather' more specifically is colexified, and this may be the relevant fact underlying the association). Guaraní and Mandarin colexify 'sky' and 'day' (see Buck 1949: 53 for Indo-European parallels), and there is a semianalyzable term in Upper Chehalis. In Buli, a term which may refer both to the 'sun' and the 'sky' is encountered, and in two languages, terms for 'sky' are based on 'sun.' These are Rama (*núnik kás* 'sun/day meat,' also colexifying 'cloud,' as noted above) and Tsafiki (*yo quido* 'sun skin'); moreover, Kapingamarangi has a semianalyzable term. Finally, Katcha and Manange colexify 'sky' with 'rain.'

Other associations include: Hausa *sama* also colexifies 'aloft' inter alia, *samaniya* dialectally also "[t]he rustle of leaves in the wind," and *gari* also 'flour, powder' as well as 'town, township,' while Khoekhoe *!āǂuisab* is related to *!ā* 'hang (laundry) out,' 'spread out to dry,' and Swahili *anga* is derived by a zero noun-class prefix from *anga* 'light.' Baruya *sigunya* also means "fat, as of a pig," Nunggubuyu *-mala-* also means 'navel' and 'thick honey,' and *yaḻamara* is also the name of the "ordinary (short-horned) grasshopper" (the meaning 'sky' is rare). Rotokas *vuvui ua* is analyzable as 'transparent CLASS.NARROW.OBJECT.' The Toaripi term *kauri* is also used to refer to a tree which "has pretty sky blue flowers," while Waris *óv* also means 'to speak, for animals to make their characteristic noise.' Yir Yoront uses 'thigh' to conceptualize 'sky': *larr-kumn* is analyzable as 'place-thigh.' The term also means 'clan.' Abzakh Adyghe *we* also means 'to burst, explode' inter alia, and Basque *ortzi* also 'space' and 'storm.' Khalkha colexifies 'sky' with 'atmosphere' and Itzaj with 'to learn,' while Nuuchahnulth has a lexical suffix colexifying "in the sky" with "on a raised platform." The Pawnee term for sky, *awaahaksuʔ,* is analyzable as /awaahak-his-uʔ/ 'be.an.expanse-PERF-NOM,' and the Xicotepec de Juárez Totonac term *a'kapūn* simultaneously denotes the 'palate.' The Bora word *íévehóówa* consists of *íéve* 'empty' and the classifier for doorways *-ʔo:gwa*. Bororo *baru* also means 'beginning,' while *waru* also means 'heat.' Cayapa *selu* also denotes a type of wave, Embera *baxã* appears to be derived from *ba* 'lightning' by means of the suffix *-xã* for covering surfaces. Guaraní *ára* is also the term for 'light' and 'time,' wheras Macaguán *bóktsebí* contains *tsebí* 'black.' Kaingang *kahnkã* also means 'family,' and the Sáliba term *mumasẽxẽ* contains *sẽxẽ* 'earth.' Tsafiki *Diósichi to* contains *to* 'land' (with *Diósichi* related to Span. *dios*?), Hani *aoq* also means 'yes, all right, okay,' 'to hold in mouth,' 'to sell' and is an interjection ('oh!'), Hawaiian *lani* is also the name of a very high chief and a kind of flower inter alia, and Kapingamarangi *langi* also means 'orgasm' as well as 'to commence.' Rotuman *ḷagi* also means 'wind' (alongside "to what purpose, wherefore"), and the Samoan term *vā-nimo-nimo* is analyzable as 'DIST-vanish-RED.'

53. The Smoke

Representation: 95%
Motivated: 33.9%
Thereof Analyzable: 7.0% Thereof Colexifying: 26.9%
Thereof by Contiguity: 6.8% Thereof by Similarity: 23.6%
Recurrent associated meanings: steam, fog, dust, cloud, fire, cigarette/tobacco, soot, spray, smell, air

Associations with other aerosols, namely, in decreasing order, with 'steam,' 'fog,' and 'cloud' are most frequent, not taking into account the obvious contiguity anchoring with 'fire.' Associations with 'steam' are mostly by colexification, with the exception of Piro and Tetun (*tš̲it̲š̲i-phya* and *ahi-suar* 'fire-vapor'). Among the colexifying languages are Buli, Efik (by the analyzable term *nsuñ'ikañ* /n-suñ-ikañ'/ 'soft/gentle-fire'), Koyraboro Senni, Burarra, Kwoma, Mali (where the relevant term is seminanalyzable, the identifiable constituent meaning 'liquid'), Ngaanyatjarra, Nunggubuyu, Sahu, Yir Yoront, Abzakh Adyghe, Upper Chehalis, San Mateo del Mar Huave, Central Yup'ik, Arabela, Bora, Carib, Cayapa (by the term *ñivijcha*, perhaps containing *ñi* 'fire, flame, seed' and *vijcha* 'difference in height'), Guaraní, Hupda, Lengua, Miskito, Ancash and Imbabura Quechua, Tehuelche, Tsafiki, Wayampi (by the analyzable term *atasĩ* /atã-sĩ/ 'fire-whiteness'), Fijian, Lenakel, and Samoan. Compare also Embera *kouwá*, meaning 'hot flash' with neuter gender alongside 'smoke' with masculine gender with *koúwa* 'vapor, fume, stink.' Buck (1949: 73) reports the pattern in Indo-European. In sixteen languages, Efik (by the analyzable term *nsuñ'ikañ* /n-suñ-ikañ'/ 'soft/gentle-fire'), Buin, Burarra, Gurindji (colexifying 'smoke-haze' with 'fog' more specifically), Yir Yoront, Abzakh Adyghe, Nez Perce (by a lexical affix), Wintu, Central Yup'ik, Hupda, Jarawara, Lengua, Maxakalí, Miskito, Rama, Wayampi (by the same analyzable term mentioned above), and Mandarin, 'smoke' and 'fog' are colexified (Buin colexifies 'white smoke' more specifically, in Rama there is also colexification with 'dew'), and in nine sampled languages, Buin, Nez Perce (again by the lexical affix), Arabela, Bora, Cavineña, Cayapa (the relevant term again being *ñivijcha* mentioned above), Maxakalí, Tsafiki and Sedang (here by the analyzable term *kia hia* 'ghost light.weight'), 'smoke' and 'cloud' are (in Buin 'white cloud' and 'white smoke' more specifically, as well as 'to be smoking tobacco' and 'be affected by smoke;' note also the similarity between Koyraboro Senni *dullu* 'smoke,' 'steam,' and *duule ~ dulla* 'cloud' as well as that the connection may be etymologically detectable for Chukchi). In some languages of the sample, more than one of the aforementioned meanings are expressed by the same term, see § 6.2.2.2. for discussion. Perhaps more surprisingly, nine sampled languages, Bezhta, Wintu (where one of the relevant term also means "cure with smoke, disinfect"), Carib, Hupda, Ancash Quechua, Fijian, Hawaiian, Sedang, and Bislama colexify 'smoke' with 'dust' (Central Yup'ik with 'dust in air' more specifically, and Ancash Quechua also with 'gas;' compare also the colexification of 'smoke' with "to be dust-windy" in Kiowa). By contiguity, 'smoke' and 'soot' are colexified in four sampled languages, namely Abipón, Xicotepec de Juárez Totonac, Ancash Quechua, and Toba (note that Itzaj *b'utz'* also means "blackened with soot"). Also by contiguity, 'smoke' and 'smell' are colexified in Kwoma and Cavineña (and note the evidence from Embera mentioned above). Six sampled languages colexify 'smoke' with 'cigarette' and/or 'tobacco:' these are Ngaanyatjarra, Abzakh Adyghe, Nuuchahnulth, Quileute, Mandarin, and Bislama (Bislama also with verbal 'to smoke a cigarette'); the Oneida term may contain a constituent with the meaning 'tobacco' as well. Sedang colexifies 'smoke' with 'air,' while in White Hmong, 'smoke' is *pa taws* 'air fire.' As seen in some of the analyzable terms mentioned above, 'fire' is an obvious choice as a contiguity anchor for the meaning 'smoke.' Other complex terms of this kind include Sko *rápong* /rapong/ 'fire-blow.at' and Kosarek Yale (Obakak valley dialect) *uk solom*, perhaps analyzable

as 'fire light.in.color.' Cheyenne directly colexifies 'smoke' with 'fire,' and semianalyzable terms in which the identifiable constituent means 'fire' are found in Kwoma, Toaripi, Waris, Bwe Karen, Lenakel, Kapingamarangi, and Manange. Yir Yoront *thorrqn* also has the meaning 'spray on waves,' and these meanings are also colexified in Fijian and Hawaiian, which latter also colexifies 'wisps.' Similarly, Bislama colexifies 'smoke' with "spray splashed up from falling heavy rain."

Other associations include: Efik colexifies 'smoke, steam' with 'exhalation' and 'heat, warmth,' as well as, from there on, "[a] good or bad influence supposed to be communicated by the heat or exhalation from the body of another." Baruya colexifies 'smoke' with 'belch, burp,' Buin *kumogana* 'thick white smoke' is derived from *kumogo* 'billow, be thick (of smoke),' Kyaka colexifies 'smoke, steam' with 'aura,' Lavukaleve with 'cheek,' and Muna *ghumbo* is also used to refer to 'many, huge numbers.' Yir Yoront *muw* is also used as a color term for 'gray' (and Kiliwa *?phuuy* is also glossed as "smokey-grey"). Abzakh Adyghe *-ɣ°e-* has very many meanings, alongside 'smoke, steam, fog' also 'dry, to make dry' and 'path, street,' while Badaga *oge ~hoge* is also the term for a "burning heap of rubbish," the 'atmosphere,' as well as a "gloomy state of affairs." *Keak*, the plural of Basque *ke* 'smoke,' can also refer to 'boasting, gloating.' Biloxi *kûsidi´* and *uksi´di* contain *si* 'yellow' and the nominalizing suffix *-di*. Yaqui *bwichia* also means 'worm,' and Jarawara *hasawiri/hasawiri* is also used to refer to 'ashes.' Yanomámi *wakë shi* is analyzable as 'red excrement.' Fijian *kuvu* also denotes "the foam at the front of a swiftly moving canoe," and Kapingamarangi *huiahi* also means "to chase away, to cause to flee." Finally, Sedang colexifies 'smoke' with "a blanket or shawl worn on shoulder for carrying child," and Yay colexifies 'smoke' and 'previous.'

54. The Soil

Representation: 80%
Motivated: 72.4%
Thereof Analyzable: 6.4% Thereof Colexifying: 66.2%
Thereof by Contiguity: 47.1% Thereof by Similarity: 0.4%
Recurrent associated meanings: land/ground, dirt, world, place, dust, clay, sand, floor, year/time, property/estate, surface, mud, field, bottom/below, day, ashes, black

For the meaning 'soil,' intra-domain associations are dominant. Most common is that with 'land' (sometimes also with 'country' as a political or administrative entity) and/or 'ground,' occurring by colexification in 67 languages, namely Efik (where the relevant term *i'söñ* is analyzable as /i-söñ/ 'NMLZ-be.hard/be.firm'), Hausa (here, the relevant term also denotes a "small, red, malodorous ant," and is also the generic name for 'snake' inter alia), Khoekhoe, Dongolese Nubian, Rendille, Yoruba, Baruya (where the relevant term also means 'ladder,' and conveys the notion of "shooting short of a target"), Buin, Burarra, Kaluli, Kwoma, Kyaka, Meyah (which also has a formally redundant analyzable term, namely *mebí efení* 'ground/soil reflection'), Muna (also colexifying 'island'), Ngaanyatjarra, Nunggubuyu, Rotokas, Sahu, Sko, Sentani, Tasmanian (all varieties except the Northern one), Waris, Kosarek Yale, Yir Yoront, Badaga, Basque, Chukchi, Greek, Ket, Khalkha (with

extension to many other related meanings), Laz, Nivkh (where the relevant term also means 'edge'), Sora, Welsh, Biloxi, Highland Chontal, Haida, San Mateo del Mar Huave, Lesser Antillean Creole French, Nez Perce, Pipil, Quileute, Xicotepec de Juárez Totonac, Yaqui, Central Yup'ik (also meaning 'village'), San Lucas Quiaviní Zapotec, Copainalá Zoque, Aguaruna, Aymara, Bororo, Carib, Cavineña, Cubeo, Embera, Huambisa, Miskito, Piro, Tsafiki, Rama, Bislama, Hawaiian, Bwe Karen, Lenakel (also colexifying 'homeland'), Malagasy, Samoan (by the analyzable terms *ʻeleʻele*, reduplicated from *ʻele* "compact red soil or stone, rust" and *palapala*, reduplicated from *pala* 'rotten'), Tetun (by the analyzable term *rai-laran* 'earth-interior'), Vietnamese (where the relevant term also means 'expensive'), and Yay. Lavukaleve has *ararume*, presumably containing *araru* 'ground,' and semianalyzable terms where one of the constituents is 'land' are found in Kaluli and Lenakel. Kyaka, Yir Yoront, Bezhta, Ket, Khalkha, Nivkh, Ineseño Chumash, Central Yup'ik, and Miskito furthermore colexify 'soil' with 'place' or 'site.'

Colexification with 'dirt' is attested in Baruya, Buin, Gurindji, Ngaanyatjarra, Nunggubuyu, Meyah, Tasmanian (Northeastern), Waris, Yir Yoront, Ineseño Chumash, Comanche, Oneida, Pipil, Tuscarora, Wintu, San Lucas Quiaviní Zapotec, Jarawara, Hawaiian (also with 'rubbish,' 'silt,' and 'excrement'), Lenakel (the term contains *tɨn* 'land'), Rotuman, Samoan, Tetun, and Bislama (22 languages). Moreover, there is a semianalyzable term where the identifiable constituent means 'earth, dirt' in Lakhota.

Another recurring pattern is the association with 'dust,' occurring in eight of the sampled languages, Efik, Yoruba, Dadibi, Ngaanyatjarra, Northeastern and Southeastern Tasmanian, Oneida, Wintu, and Toba (and evidenced by cognates from Old English and Gothic, one meaning 'soil,' the other 'dust,' Buck 1949: 18). Seven further sampled languages colexify 'soil' with 'clay,' namely Buli, Efik, Baruya, Bezhta, Japanese, Hupda (also colexifying 'Tuyuca people'), and Bislama, and, similarly, in Koyraboro Senni, San Mateo del Mar Huave, and Samoan, the same term is used for both 'soil' and 'mud.' Bezhta and Cubeo colexify 'soil' with 'field,' and Lesser Antillean Creole French, San Lucas Quiaviní Zapotec, and Mandarin with 'property, estate.'

In six sampled languages, Efik (by the same analyzable term mentioned above), Basque, Guaraní, Macaguán, Imbabura Quechua, and Tetun, the term for 'soil' is also used for the 'floor' (e.g. of a house), and in another six, Buli, Efik, Rendille, Ngaanyatjarra, Tehuelche and Rotuman, 'soil' and 'sand' are colexified (the Buli term is semianalyzable, containing *tain* 'stone, pebble,' and the Tehuelche term also means 'sand dune'). Kapingamarangi has the corresponding analyzable term *gelegele luuli* 'sand black' (compare also Tetun *rai-metan* 'earth-black'). Oneida and Wintu colexify 'soil' also with 'ashes;' the relevant Wintu term *bukul* is related to *buk* 'dark.'

There are also associations which move to more abstract spatial relations. For instance, the Efik term mentioned above, as well as an unanalyzable Hausa term also mean 'bottom, below' (see Heine and Kuteva 2002: 121 generally and Buck 1949: 18 for a parallel from Latin; note also that Embera *udáa* has the adjectival meaning 'down' alongside nominal 'soil, ground'), and also in Efik, as well as in Abzakh Adyghe and Bora, terms bearing an association with 'surface' in general occur: Abzakh Adyghe *sʔəgʷ-ɕẖeɕʷə* is analyzable as 'earth-surface,' and Bora *ííñújɨ hallu* consists of *ííñú* 'earth' suffixed with the classifer for

disc-like objects *-jɨ* and *hallu* 'top, outside part.' In Efik, Rendille, Abzakh Adyghe, Basque, Ineseño Chumash, Quileute, San Lucas Quiaviní Zapotec, Aguaruna, Carib, Guaraní, Huambisa, and Ancash Quechua, as in other European languages, terms for 'soil' may also refer to the entire 'world.'

Even more interesting are extensions into the temporal domain, that is, terms for 'soil' than can also refer to a 'year' and/or 'time' in general. Kyaka, Ket, Highland Chontal, Ineseño Chumash, and Ancash Quechua are languages in the sample with terms that behave like this, while Kyaka and Yir Yoront colexify 'soil' with 'day' (in fact, *larr*, the relevant Yie Yoront term has a very wide semantic range); Holmer (1966) also reports similar patterns of colexification for languages of Oceania. Kyaka furthermore colexifies "event, opportunity, chance" as well as 'weather.'

Other associations are: Efik *isöñ* colexifies 'soil' with "hatch or trap door," *umabu* colexifies 'soil' with 'mould,' and *n'tan* is also the name of a plant with flowers that eject a dusty substance when being touched. Toaripi *mea* colexifies 'soil' with 'wind, weather' inter alia, while Sahu *tana'a unisi* "dry, infertile soil on a ridge" is analyzable as 'earth/ground/land shin/shin.bone.' Kosarek Yale *soko*, dialectally used with the meaning 'earth, soil' also denotes a 'special type of tie rod' without dialectal restrictions. Abzakh Adyghe *sʔə* also means 'to produce, to construct' inter alia, and *sʔəg°*, a complex term of the redundant kind, can also refer to the 'ground floor.' Badaga *parava* means "bothered, concerned, troubled" and is also the name of a specific kind of soil, and Khalkha *kørysyn ~ kørydesyn* also means 'crust, peel, rind.' Welsh *gweryd* is also used as a term for 'grave.' The root *-ir-* yielding Tuscarora *à·wiʔr* 'soil' can also refer to a "bit, grain, particle" and 'small piece,' Kaingang *ga* also means 'louse, worm,' and Piro *tš̲i-xi* appears to be a diminutive of *tš̲i* 'fire, firewood.' Rama colexifies 'earth, ground' with 'going,' Fijian with 'cluster, shoal, swarm,' and Bwe Karen *ha* also means 'hole in the ground, pit' inter alia. Lenakel *nɨmɨtɨk* 'red or reddish soil' might contain *nɨmɨt* 'mud, swamp,' Hawaiian *lepo* figuratively can also refer to 'common people,' Samoan *ʻeleʻele* (see above for morphological analysis) also means 'blood' and 'menses' in polite usage, while Tetun *rai-metan* 'earth-black' colexifies 'fertilizer.'

55. The Spark

Representation: 54%
Motivated: 39.1%
Thereof Analyzable: 27.2% Thereof Colexifying: 13.1%
Thereof by Contiguity: 10.0% Thereof by Similarity: 23.1%
Recurrent associated meanings: fire, lightning, embers, flame, burst/explode, light, particle, sparkle, firefly, flower, star, grain

There are a number of lexico-semantic ties pertaining to the meaning 'spark,' but none of them is particularly frequent. Basque, Khalkha, Haida, and Tuscarora colexify 'spark' with 'embers' (Tuscarora in addition colexifies 'candle,' 'flash of light,' 'light,' 'lamp,' and 'taper'). In three languages, the term for 'spark' is associated lexically with verbs meaning 'to burst, explode.' These are Nivkh (*p'ryrk t'uɣr* 'burst fire'), Nez Perce (*taxliċáˑsa*, containing

taẋ 'explode' and *liċé·* 'be directed at'), and Fijian (*lidi ni buka* 'burst/explode POSS fire/firewood'). The same word is used for 'spark' and 'flame' in four sampled languages, Abzakh Adyghe, San Mateo del Mar Huave, Bororo, and Chayahuita (the Abzakh Adyghe and Chayahuita terms in fact are complex containing the words for 'tongue;' since this is a very frequent pattern in terms for 'flame' as described in section 22, it seems reasonable to assume that this is indeed the primary meaning in these languages). Tuscarora and Central Yup'ik have associations between 'spark' and 'light' (Central Yup'ik by the term *kenurraq*, which is perhaps analyzable as /keneq-rraq/ 'fire-little.bit,' the meaning 'spark' is attested only for the dialect of Norton Sound, in other dialects it means 'light' or 'lamp'). Relatedly, in Khoekhoe, *nanib* ~ *nanis* is derived from the verb *nani* 'to twinkle, flicker, gleam, burn slowly,' and there is a semianalyzable term in Kaingang. Hawaiian and Rotuman have complex terms in which one constituent means 'particle:' *huna-ahi* 'particle/speck/crumb/grain-fire' (also denoting 'live cinder') and *momoe ne rạhi* 'fine.particles ART.PL fire.' Also in two languages of the sample, an association with the meaning 'to sparkle' is found, by colexification in Bororo and by the term *yantsáji* /yáants-ji/ 'sparkle-POSS' in Aguaruna. Contiguity-based associations by morphologically complex terms with 'fire' making reference to the fact that sparks are, in Cognitive Linguistics parlance, a figure emerging from a larger ground structure (the fire) are Kashaya *ʔoho cuhṭʰucuhṭuw*, analyzable as /ʔoho cu-hṭʰuʔ-cuhṭʰuʔ-w/ 'fire round.object-pieces.come.off.bigger.object-RED-ABS' and Wintu *pʰo·h dil-ma* 'fire drop/fall/alight-??' (this term itself is glossed as "Sparks are flying. He dropped the fire"). There is also at least one language which directly colexifies 'fire' with 'spark,' namely Chukchi; the association may also be present in Northeastern Tasmanian.

Recurrent metaphor-based associations are also found. Abzakh Adyghe, Badaga, Abipón, Bora, Ancash Quechua, and Lesser Antillean Creole French colexify 'spark' with 'lightning.' Berik has *tokwa es*, presumably analyzable as /tokwa ese/ 'fire flower' and Tetun *ahi-fuhan*, also analyzable as 'fire-flower.' In two sampled languages, the word for 'spark' contains that for 'star': Dadibi *sia hó* 'fire star' and Bislama *sta blong faea* 'star POSS fire.' In Baruya, 'spark' is *di'nyaala* /dika-nyaala/ 'fire-firefly,' Bezhta directly colexifies the relevant meanings, while Huambisa has the semianalyzable term *yantsari*, for which compare *yantsa* 'firefly.' Miskito has *pauta yuya* 'fire grain,' and Hawaiian *huna-ahi* 'particle/speck/crumb/grain-fire.' Other metaphorical associations in which 'fire' acts as a contiguity anchor include: Yoruba *owọ́-iná* 'hand-fire,' Toaripi *a-e*, perhaps analyzable as 'fire-faeces,' Ket *bógdes*, analyzable as /boʔk-dēs/ 'fire-eye' (note that Haida *sráahld* is derived from a verb meaning 'to glance'), Xicotepec de Juárez Totonac *kosa macscut* 'jump/get.up fire.' Semianalyzable terms in which 'fire' figures are found in Efik, Toaripi, Kosarek Yale, Carrier, San Lucas Quiaviní Zapotec, and Arabela.

Other unsystematic associations include the following: Muna *wara*, when used verbally, also means 'to drickle, to drip,' and Yoruba *ẹ̀ta* also means 'splash' as well as 'root, tuber.' Ngaanyatjarra *tii* also means "healed tissue" as well as, by English influence, 'tea,' while Nunggubuyu *-ṛarwadawada-* may be related to *ṛa:r* "burnt-out grassland or light bushland" and *=wˌada-* 'to snap, to break suddenly.' The second constituent in the Yir Yoront term *thumlilqli* resembles *lilq*, which means 'alone, by oneself' (*thum* is 'fire').

Basque *pindar* also colloquially means "spunk, pep," and *txinpart* is also used with the meaning 'livewire.' Khalkha *cindara* contains or is otherwise related to *cindar*, a respectful term for 'remains, corpse' and also means "white ashes on dying coals." *Co*γ also denotes 'hot ashes,' as well as, figuratively, "splendor, grandeur, glory; energy, spirit." Welsh has *tamaid o beth llosg* 'piece of something burning,' and Blackfoot *ipáísstsitiimi'kaa* 'to throw a spark' contains *mi'k* 'red.' The denotational range of the Itzaj term *se'es* covers also "sawdust, bit, small chip, crumb, confetti" as well as "finely cut, fine." The Kiliwa term contains an element meaning 'earth.' Wappo *ṗétši?* also means "snap like burning wood," and Copainalá Zoque colexifies 'light thunder.' Arabela colexifies 'piece of lit coal,' and Bororo *beri* is also used with the meanings 'abundance' and 'arrogance,' while Embera *adyizúa* colexifies 'brilliant' and 'bright.' Great Andamanese *châpal'igbêra* contains *châpa* 'firewood' and *bêra* 'sweepings,' and Hani *miqseil* might contain *seil* 'louse' (the term also denotes a 'woman dissatisfied with her marriage'). Kapingamarangi *madagologolo* is related to *kolo* 'to drill, twist as a knob.' Samoan *sipaka* (< Engl. *spark*) also denotes a 'spark plug.'

56. The Spring

Representation: 87%
Motivated: 46.3%
Thereof Analyzable: 35.5% Thereof Colexifying: 12.0%
Thereof by Contiguity: 27.8% Thereof by Similarity: 12.6%
Recurrent associated meanings: water, river, hole, eye, come out, lake/pond, jump, head, puddle, boil, dig/dug, fetch water, headwaters

Motivated terms for 'spring' (or 'well,' which was accepted as a proxy) are frequently morphologically complex, with one of the constituents being 'water.' However, also for the meaning of the second constituent involved, there are a number of recurrent patterns. In twelve sampled languages, this is 'hole,' as in Bakueri *e̱wondí yá máliwá* 'hole of water.' Alongside Bakueri, words with this structure are attested for Berik, Toaripi, Haida, Yuki, Bora, Bororo, Guaraní, Hupda (where the word for 'hole' also means 'house'), Maxakali, Rama, Yanomámi (where an additional element meaning 'point' is present), and White Hmong (where there may be a further element meaning 'issue forth' present). In addition, Embera colexifies a general term for 'hole, pit' with 'spring' and in Hani, 'village well' is *lolhovq*, with *lol* being the classifier for rivers and *hovq*, meaning 'to fetch water' (for which association in turn compare Chukchi *ajmə-n* 'fetch.water-LOC') as well as being a classifier for pits and holes (see also Buck 1949: 45 for this semantic connection in Indo-European, evidenced in Greek and Armenian). Somewhat similarly, Cubeo has *jiacarã-cobe*, probably containing *jiacacʉ* 'be.aquatic' and a classifier for hole-like objects. The second most frequent pattern is metaphorical in nature, with the second constituent being 'eye,' as in Meyah *mei eitéij* 'water eye' (see also § 6.2.3.1. for discusion of 'eye'-metaphors). Other languages with such terms are Buli, Kyaka, Sahu, San Mateo del Mar Huave, Bislama, Fijian, and Tetun (note also Welsh *llygad ffynoon* 'eye well/fountain' and similar redundant terms in Khoekhoe, Muna, Ancash Quechua, and Hawaiian). Hausa, Dongolese Nubian, Burarra, Ancash Quechua, and Samoan colexify 'eye' with 'spring;' due to the pervasive-

ness of metaphors on the basis of 'eye' cross-linguistically (see Buck 1949: 44 for semantic development from 'eye' to 'spring' in Armenian), many of the languages have further meanings colexified. The association with 'eye' is also common in Semitic languages (e.g. Segert 1991: 1432). Another metaphorical pattern is constituted by the transfer from 'head' to 'spring,' as in Yanomámi *u he* 'liquid/river head.' This association is also found in Mbum, Sora (where an additional element meaning 'hill' is present and the relevant term denotes a "spring of water on the hill"), Malagasy, and Tetun; Miskito colexifies 'head of plant' and 'spring,' and a semianalyzable term of this kind, colexifying 'source' (also in the sense of 'source of information') and 'origin' is also found in Basque; also note Yoruba *orísun* 'head-spring.or.fountain.' In four sampled languages, Mbum, Yoruba, Ancash Quechua, and Rotuman, terms for 'spring' are encountered which contain a verb meaning 'to jump' or to 'spring,' for instance Mbum *hvíŋà-mbìì* 'spring.out water' for a 'fountain;' the relevant Fijian term is reduplicated from a verb meaning 'for water to spring up.' Samoan colexifies verbal 'jump, leap' with 'boil' and nominal 'spring, source' (for the association with boiling, note that Central Yup'ik *qalla-neq* colexifies 'spring' with 'eddy' and is analyzable as 'be.boiling-thing.that.results.from' and that the Swahili term *chemchemi* is derived by reduplication from the verb *chemka* 'to boil'). Somewhat similarly, in eight sampled languages, One, Japanese, Nivkh, Sora, Kiowa, Bora, Chayahuita, and Takia (in Japanese and Bora, there are uncertainties as to the analysis), there are at times quite complex terms for 'spring' revolving around verbs meaning 'to come out,' 'to go out,' or 'to exit,' such as One *fola suwe* 'water/river come.out;' note also that in Kiliwa *ʔmatcpam,* the sequence *-cpam* (*mat* is 'earth') might derive diachronically from **c+paa* 'come out.' Buin colexifies 'spring,' 'to emerge' and other meanings directly, and Welsh *tarddiad* is derived from *tarddu* "to spring, to sprout, to derive from, to issue." In Biloxi, perhaps Cheyenne, Lakhota, and Tehuelche, consituents meaning 'to dig (out)' or 'dug' figure, e.g. in Lakhota one of the terms for 'spring' is *mničʔápi* /mní-čʔápi/ 'water-dug' (Cheyenne colexifies 'spring with 'water pump' and 'windmill'); Rotokas and Copainalá Zoque have semianalyzable terms with such structure.

Twelve sampled languages, Bakueri, Buli, Efik, Ngambay, Muna, Sko, Badaga, Kashaya, Pipil, Maxakalí (by the complex term *kõnãgkox*, analyzable as /kõnãˈãg-kox/ 'water-hole'), Bwe Karen, and Kapingamarangi directly colexify 'spring' or 'well' with 'river, stream' by contiguity (Ngambay also with 'waves;' the connection is evidenced by a cognate set between Latvian and Sanskrit in Indo-European, Buck 1949: 45). Buli, Wintu, Yuki, Cayapa, Kaingang, and Maxakalí colexify 'spring' with 'pool, lake' and/or 'pond' (again by the analyzable term already mentioned), Similarly, in Sko, the 'spring' is called *pa-í* 'water/river-pool.' In Basque, Oneida, Pipil, and Hawaiian, 'spring, well' is colexified with 'puddle.' Due to colexification of 'water,' 'river,' and 'lake,' the association with 'lake' is also found in Berik by an analyzable term, and there is a semianalyzable term in Quileute.

This brings the discussion back full circle to the frequent role of 'water' to act as contiguity anchor in complex terms, and the different patterns of colexifications of 'water' with different bodies of water (see also § 6.2.2.5.). Ngambay, Muna, Yir Yoront, Comanche, and Kashaya directly colexify '(spring) water' and 'spring' (and sometimes also other bod-

ies of water, further, Ngambay also colexifies 'waves' and Badaga also 'dale,' 'flat land' and 'riverside').

Noni *fijoo* 'spring' consists of *joo* 'water' and the noun class prefix *fi-*, and Lavukaleve *lafio* 'spring' is connected to *lafi* 'water.' Bwe Karen has *chí-bú* 'water-in.' However, alongside those already mentioned, there are also a relatively large number of language-specific conceptualizations realized by morphologically complex terms on the basis of 'water.' These are: Noni *joo yi caan* 'water REL small,' Kaluli *ho:n sí* 'water tip' (compare the possible etymology of Ket *tájlop* < **taj-ūl-ʔqōp* 'cold-water-tip'), Kyaka *ipwua renge* 'water source' (with *renge* also having other readings), Meyah *mei ofog* 'water round,' Sahu *'banyo ma utu'u* 'water POSS root,' Abzakh Adyghe *psə-λaqʷʲe* 'water-foot/stem,' Ket *aqtul*, analyzable as /aqta-ūl/ 'good-water,' Carrier *thaḳezḳět*, containing *tha* 'water' and *ḳět* 'place,' Oneida *kahnekóniʔ* /ka-hnek-No-ʔ/ 'NEUT.AGENT-liquid/liqour-be.in.water/cook.in.water-STAT,' Pipil (Santo Domingo de Guzmán dialect) *puːni aː-t* 'be.born water-ABS,' Santiago Mexquititlan Otomí *pothe* /poho-dehe/ 'well.up-water,' Wintu *poˑpil-mem* 'summer-water/river' (also denoting a "spring that dries in the winter"), Yuki *pil-ʔuʔuk* 'snow-water,' Guaraní *y-vu* 'water swollen,' Wayampi *ɨapɨ* /ɨ̈-apɨ/ 'water/river-source,' colexifying 'valley' and 'dew,' Bislama *maot blong wota* 'mouth POSS water,' and Hawaiian *kumu wai* 'foundation/base water' and *wai hū* 'swell water' (meaning "gushing spring, overflowing water"). There are semianalyzable terms with 'water' in Kosarek Yale, Comanche, Quileute, Yana, Yuki, Bislama, and Kapingamarangi. Baruya and Kosarek Yale colexify 'spring' with 'headwaters,' and similarly, Sedang colexifies 'upstream.'

Other associations include: the Khoekhoe terms *laub* and *laus* contain the root *lau* "trickle, purl, run/flow gently," and Rendille *wór* also means 'news.' Anggor *fe amoŋgo* seems to contain *amoŋgo* 'sibling,' Basque colexifies 'well' and 'puddle,' while Sahu *gogonyo'o* contains *onyo'o* 'to draw water.' Khalkha *bulaɣ* also means "[h]aving white spots, partly white...," and Welsh *ffynhonnell* is derived from *ffynhonni* 'to well, to gush.' Ineseño Chumash *'aqmilimu'* is derived from the verb *'aqmil-* 'to drink,' while Itzaj colexifies 'spring' with 'splash' and Lake Miwok *ʔóla* is also a kinship term. The Pawnee term *kicaahkatakus* is analyzable as /kic-haahka-ta-kus/ 'be.liquid-be.attached-suspended-be.sitting,' Santiago Mexquititlan Otomí colexifies 'spring' with 'black,' and the Tuscarora term *učaʔtuhstaʔkyéhaʔ* is based on the verb root *-čaʔtuhsT-* 'be cool.' Guaraní *ykuayvu* also denotes an 'underground watercourse.' Hawaiian *hāpuna* colexifies 'spring' with 'coral,' 'lime' and other things, and *māpuna* "bubbling spring" with "froth, as of a rough sea" and "surging of emotions." Mandarin colexifies 'spring' and 'neck' (the relevant lexical items were still distinct in Early Middle Chinese though), and Rotuman colexifies 'water source' with 'medicine,' 'cask,' and "stew or hash made of meat or fish."

57. The Star

Representation: 97%
Motivated: 21.2%
Thereof Analyzable: 5.2% Thereof Colexifying: 18.1%
Thereof by Contiguity: 2.7% Thereof by Similarity: 19.2%

Recurrent associated meanings: planet, firefly, starfish, meteoroid, moon, constellation, blaze, shine/sparkle/blink, asterisk, badge of rank, dot/spot, fire

'Star' is a meaning expressed in many languages by an unanalyzable, monomorphemic word. Semantic associations by colexification are also relatively rare. Muna, Nunggubuyu, Rotokas, Toaripi, Tuscarora, Bororo, Wayampi, Yanomámi, Fijian (where the relevant term *kalokalo* also denotes the flower 'aster,' the English name of which incidentally itself goes back to the Ancient Greek word for 'star;' *kalo* itself is 'pull bowstring, discharge gun'), and Sedang colexify 'star' with 'planet' (Sora and Toaripi also with 'comet' or 'meteor;' moreover, in Khoekhoe, the same term suffixed with different nominal designants yields the meanings 'star' and 'comet' respectively, while Bislama *sta* is glossed as "any heavenly body (e.g. moon, star, meteorite)"). The second most common association, found in Buin, Muna (where the term *kolipopo ~ ngkolipopo* may be related to *popo* "evil spirit which looks like a flashlight attacking people"), One, Waris, Kosarek Yale, San Lucas Quiaviní Zapotec, Bora, Cavineña, and Yanomámi, is colexification of 'star' with 'firefly' by a metaphorical transfer based on perceptual similarity; the relevant Yanomámi term is semianalyzable containing an element meaning 'round fruit' and also denotes an unidentified species of oruga.

Otherwise, there is scattered evidence for lexico-semantic ties between 'star' and the two large heavenly bodies, the 'sun' and the 'moon.' Burarra is the only language in the sample which colexifies 'sun' and 'star' (and concomitantly, a type of shellfish similar in appearance to a star, several species of sea urchin, and further meanings associated with 'sun,' such as 'watch,' see section 60), though note that Burarra also features an unrelated monomorphemic term for 'star' specifically. Semianalyzable terms for 'star' in which one constituent appears to be 'sun' are found in Maxakalí and White Hmong, and in addition, Hupda *wædhɔm'ɑ̈h* might be analyzable as /wædhɔ-mæh/ 'sun/moon-small,' although this is not entirely straightforward. The source remarks that the words for 'star' are also based on a word for 'sun, moon' in languages of the neighboring Tucanoan family. Ties with 'moon' specifically also exist. Next to the case of Bislama already mentioned, Abipón is the only language in the sample colexifying the two by the term *eergRaik*, derived from *eerg-* 'to burn, sparkle.' In the two sampled Tupi-Guaraní languages, Guaraní and Wayampi, the word for 'star' is analyzable as 'fire-moon' (*jasyrata* /jasy-tata/ and *yaitata* respectively; Guaraní also has the variant *yvágarata* /yvaga-tata/ 'sky-fire'), and in addition, there are a number of languages in the sample where a diachronic association may exist. These are suspiciously concentrated in Africa: the Buli word for 'star,' *chingmarik*, might contain *chiik* 'moon' and *ngmari* 'take from' (this analysis is marked as questionable in the source). Likewise the analysis of Efik *ntan'tafiöñ* is dubious: it might be a complex term consisting of *n'tan* 'earth, dust' and *ö'fiöñ* 'moon.' Koyraboro Senni is yet another African language with a term for 'star' that is apparently related, at least in a diachronic sense, to other lexical elements: *handarey* resembles both *handi* 'day' and *handu* 'moon, month.' For Dongolese Nubian *wíss(ɪ)*, Armbruster (1965) suggests an etymology connecting the term to a word for 'moon' plus a diminutive suffix; note also the similarity between Rendille *yeyyehím* 'star' and *yéyyaH* 'moon.'

Moreover, Carib, Tehuelche, and Wayampi use the same term for a 'star' and a 'constellation of stars,' and Toaripi, Sora, and Bislama employ a single term for both 'star' and 'shooting star, meteor.'

Another class of terms for 'star' are those derived from verbs meaning 'to shine,' 'to sparkle,' 'to blink' etc. (see Buck 1949: 56 for the association with 'shine' in Sanskrit). As already mentioned, Abipón *eergRa-ik* is derived from a verb meaning 'to sparkle' and similarly, in Khoekhoe, the word for 'star,' *lgami-ro-s*, is analyzable as 'blink-DIM-3SG.FEM' (a variant is *lhom-lgami-ro-s* 'sky-blink-DIM-3SG.FEM'). In Hani, the 'star' is called *aqgeel alsiq* or *aqgeelsiq*, with *geel* meaning 'to shine' and *alsiq* 'fruit;' *siq* also acts as a classifer for round things, inter alia (for 'fruit,' compare Austin et al. 1976: 61, table 2 for evidence from Arabana and Waŋgaŋuru), and Piro *katahiri* apparently contains *kata* 'shinging, blazing.' Three sampled languages, Hausa, Basque, and Kildin Saami, colexify 'star' with 'blaze;' the Basque term also may refer to a 'cataract,' 'asterisk' (an association it shares with Lesser Antillean Creole French), 'dream,' and a 'star' in the sense of a celebrity as well as someone 'dear, beloved.'

Other recurrent associations include: In four languages of Oceania, Buin, Lavukaleve, Rotokas, and Bislama, 'star' and 'starfish' are designated by the same term. In two sampled languages of Eurasia, Basque and Khalkha, 'star' is extended to 'badge of rank' (a pattern also found in English, German, and presumably other languages of the region). Kyaka and Oneida colexify 'star' with 'dot' and/or 'spot' (Oneida also with 'print').

Still other associations include: Buin *kaipa* may also refer to "anything star-shaped," Gurindji *kiki* also to a type of ornament, while Meyah *motúr* is semianalyzable: it contains *motú* 'night.' One *leila* also denotes a river frog, Sko *ha* also means 'bag' and 'walk,' Kosarek Yale has a semianalyzable term containing an element meaning 'sky, air,' Waris *pai* is also an interjection ("my!"), Badaga *mi:nu* also means 'fish, shellfish,' Bezhta *cã* also 'salt,' and Kolyma Yukaghir *jurgud'e:jə ~ jurgud'e:jjə*, colexifying 'awl,' contains *jurgu:* 'slot, hole' (Tundra Yukaghir has *paɣad'iid-ekuu* 'drill-hole,' Nikolaeva 2006: 340). Ineseño Chumash *'aqiwo* also means 'snail.' The Haida term *k'a7ihldaa* is analyzable as /k'a-7ahlda/ 'tiny.object-glance.at,' Itzaj colexifies 'star' with 'black,' and the Kashaya term *q^{h}a·moṡ* contains *moṡ* 'sour.' The first element may be etymologically related to *q^{h}a·ʔa* 'nightlong' or *q^{h}aʔaw* 'morning.' Tuscarora *uʔnihsę·reh* also means 'navel.' Wintu *λuyu·q* is related to *λu* 'to stab' and also denotes 'porcupine, porcupine needles.' Arabela *rijia* also means 'earthquake.' Chayahuita *tayora* probably contains the classifier *-ra* for 'small things,' and Lengua *yoa* also denotes a 'pebble.' Miskito *karma* also means 'throat' and 'origin.' Rama *piúp* contains *up* 'eye,' and Yanomámi *kurikayari* might be related to *kurikaya*, a term for a parrot species. Bwe Karen colexifies 'star' with 'to run' inter alia, and Rotuman *sina* also means 'light, lamp' inter alia (for this term the lexicographer remarks that it might be restricted with the meaning 'star' to a single fixed expression). The Yay term for 'star,' *daaw^{1} di^{2}*, might contain *daaw^{1}* 'to stir,' and Vietnamese *sao* also means 'how.' Lesser Antillean Creole French *étwal* also means 'destiny.'

58. The Steam

Representation: 70%
Motivated: 53.9%
Thereof Analyzable: 11.3% Thereof Colexifying: 42.7%
Thereof by Contiguity: 17.5% Thereof by Similarity: 29.4%
Recurrent associated meanings: smoke, fog/mist, breath/exhalation, heat/hot, smell, cloud, water, dust, air, gas, boil, spray, sweat, fire

The meaning 'steam' (for which 'vapor' was accepted as a proxy) is frequently associated with other dispersions of particles in the air, such as 'smoke,' 'cloud,' and 'fog.' 31 languages colexify 'steam' with 'smoke;' these are Buli, Efik, Koyraboro Senni, Burarra, Kwoma, Mali (the relevant term *chulēski* contains *chulēs* 'liquid'), Ngaanyatjarra, Nunggubuyu, Sahu, Yir Yoront, Abzakh Adyghe, Upper Chehalis, San Mateo del Mar Huave, Central Yup'ik, Arabela, Bora, Carib, Cayapa (by the term *ñivijcha*, perhaps containing *ñi* 'fire, flame, seed' and *vijcha* 'difference in height), Guaraní, Hupda, Lengua, Miskito, Ancash and Imbabura Quechua, Tehuelche, Tsafiki, Wayampi, Fijian, Lenakel, and Samoan. Morphologically complex terms are also found: in Highland Chontal, 'steam' is *liguxís gajah* /liguxís lajah/ 'smoke water,' and in Santiago Mexquititlan Otomí *'bipa* is analyzable as /'bifi-pa/ 'smoke-heat.' In Hawaiian, one term for 'steam' is *uahi wai*, where *uahi ~ uwahi* means 'smoke, spray' and *wai* 'water, liquid' (compare also the colexification of 'steam' with 'spray from waves' in Yir Yoront and with 'spray' generally in Fijian). Similarly, Lake Miwok *ṣímúuṭi* also means 'for smoke to emerge.'

In addition, in two of the languages with terms colexifying 'smoke' with 'steam,' they are morphologically complex, and their internal structure suggests that 'smoke' is the dominant meaning. Efik *nsuñ'ikañ* contains *suñ* 'soft gentle' and *ikañ'* 'fire,' and in Wayampi, *atãsĩ* is analyzable as /ata-sĩ/ 'fire-whiteness;' moreover, the relevant Kwoma term is semianalyzable, with the identifiable constituent meaning 'fire.' However, the direction of mapping as revealed by the evidence from complex terms is not entirely unidirectional, as evidence from Tetun shows (see section 53).

Sixteen languages in the sample, namely Efik, Anggor, Burarra, Kyaka, Yir Yoront, Abzakh Adyghe, Greek, Khalkha, Welsh, Wintu, Arabela, Hupda, Lengua, Miskito, Hawaiian, and Rotuman colexify 'steam' with 'fog, mist' (see Buck 1949: 66 for the connection with 'vapor' in Indo-European), and five, Anggor, Arabela, Bora, Cayapa (by the analyzable term mentioned above), and Tsafiki colexify 'steam' with 'cloud' (note also the similarity between Koyraboro Senni *duule ~ duula* 'cloud' and *dullu* 'smoke, steam'). Moreover, four sampled languages, Carib, Hupda, Ancash Quechua, and Fijian colexify 'steam' with 'dust,' and in Wintu, Ancash Quechua, and Hani, the same term is used for 'steam' and 'gas' (the relevant Wintu root *ẋOs* also yields *ẋoso* 'lungs'). As is always the case with the associations pertaining to aerosols, it should be borne in mind that some sampled languages use the same term for not only two, but sometimes three or even four of these meanings (see also § 6.2.2.2.).

An association that is particular to the meaning 'steam' is that with 'breath, exhalation,' occurring by colexification in Bakueri, Efik, Buin (where the term also denotes the

'windpipe, trachea' as well as "Puffing, breathlessness; airiness"), Kosarek Yale, Bezhta, Khalkha, San Lucas Quiaviní Zapotec, Ancash Quechua, Fijian, Hani, and Sedang (inter alia), and by the analyzable term *sʜdl-hą̄'-t'ǫ̨-gyʜ* 'be.hot-breath-water-NOUN.POSTFIX' in Kiowa. As in this language, analyzable terms for 'steam' frequently contain an element referring to 'heat' (and Efik, Yoruba, Itzaj, and Pawnee, by the analyzable term *awiriitu?* /awirit-u?/ 'be.hot-NOM,' colexify the meanings directly, while Bislama colexifies 'steam' with "radiated heat from sea" more specifically). Santiago Mexquititlan Otomí has, as already mentioned above, *'bipa*, analyzable as /'bifi-pa/ 'smoke-heat,' Yanomámi has *u heõshi* 'liquid hot,' and Takia *you wanna-n san ur* 'water hot-3SG its air' (for the association with 'air,' note also Vietnamese *hơi nước* 'air water' and the colexification of these meanings in Bakueri and Ket alongside the colexification with 'atmosphere' in Khalkha). A semianalyzable term where one constituent can be identified to mean 'hot' is found in Rotuman. Completetly unrelated patterns are the colexfication of 'steam' with 'smell, odor' and/or 'fragrance (from cooking),' which is found in Ngambay, Burarra (here the relevant term *-jinyja* is related to the verb *jinyja*, meaning 'be standing, be vertical' inter alia), Gurindji, Kwoma, Basque (where the term also denotes 'perfume' specifically), Bezhta, Tuscarora (by an analyzable term containing the roots *-hsęri-* 'savor' and *-ur-* 'to cover'), and Yay. Colexification of 'steam' and '(to) sweat' is found in Santiago Mexquitilan Otomí (by the analyzable term *xa̱ni-dehe* 'to sprinkle-water') and White Hmong. As has emerged from the discussion so far, 'water,' associated by contiguity with 'steam,' frequently figures in analyzable terms. Another such term not yet mentioned is Santiago Mexquitilan Otomí *ho̱e-dehe* 'to.fall-water,' and semianalyzable terms where one of the constituents is 'water' are also found in Kiliwa and Sko alongside Japanese, where a term for 'steam' contains *yu* 'hot water' (see Goddard 2001 on the lexical distinction between hot and cold water in Japanese). Abipón has *l-apa-Ra*, analyzable as 'POSS.INDEF/3SG-boil-ABSTR' and colexifying 'foam,' Great Andamanese *bôag* has a verbal reading as 'to boil,' and a verb colexifying 'to steam' with 'to boil' is present in the relevant term in Oneida.

Other associations include: Efik *uye* also denotes "[a] good or bad influence supposed to be communicated by the heat or exhalation from the body of another," and Muna *oho* also means 'to feed' as well as 'close lid, cover.' Kosarek Yale *iba* metaphorically also means 'fury, passion.' Abzakh Adyghe *-ɣ°e-* also means 'to dry, to make dry' and 'path, street' among many other things, Badaga *a:vi* also means 'yawn, yawning,' as well as 'spirit, soul' and related notions, Khalkha *aɣur* metaphorically also means 'anger,' while Welsh *anwedd* also means 'enormous' and other things. Cahuilla *mululiš* is derived from the verb *-múlul-* 'to come out steaming, bubbling' (similarly, Nez Perce *mú·yn* is derived from *mú·y-* 'to rise as steam,' and Chayahuita *tomontërinso'* from *tomoitërin* 'for steam to rise up'). Carrier *yentseł* "steam (over the ground after rain)" contains *yen* 'earth,' Itzaj colexifies 'scab' inter alia, while Santiago Mexquititlan Otomí *h-ñä* is analyzable as 'IMPERSONAL.VOICE-speak.' The term means 'steam' as well as 'voice, language.' Wintu *λi·d* also means to "extinguish fire, sprinkle water on hot rocks to make steam, throw water on fire" and *x0s* also means 'lungs' (which is suggestive of an underlying association with 'breath'), Chayahuita colexifies 'steam' with 'airplane,' Guaraní with 'nose, beak, point,' and Hani with 'to cook by steaming.' Bwe Karen *θu* also means 'blood' inter alia, Fijian *kuvu* also denotes "the

foam at the front of a swiftly moving canoe," and Hawaiian *māhu* also several kinds of trees.

59. The Straw

Representation: 44%
Motivated: 70.0%
Thereof Analyzable: 26.2% Thereof Colexifying: 43.8%
Thereof by Contiguity: 53.3% Thereof by Similarity: 6.9%
Recurrent associated meanings: grass, dry, maize, chaff, stalk/cane, fodder, pasture/lawn, mat, kindling, hat

By far the most frequent association for 'straw' (or 'hay') is that with 'grass, weed,' either by colexification (sometimes with additional colexified meanings, see section 28), or by morphologically complex terms, which have, in the overwhelming majority of cases, an element meaning 'dry' as their second constituent. The association is also common in diachrony in Indo-European (Buck 1949: 520-521). Colexification is found in Buli, Ngambay, Abzakh Adyghe, Badaga, Bezhta, Khalkha, Kildin Saami, Upper Chehalis (by the term *s-máqʷm=umš* 'CONTINUATIVE-prairie=place'), Cheyenne, Chickasaw, Highland Chontal, Itzaj, Kashaya, Kiowa, Lakhota, Nez Perce (also colexifiying 'lettuce' in one of the relevant terms and "bits of straw to start fire" and 'kindling' in the other, the latter association is shared with Ancash Quechua), Nuuchahnulth, Oneida, Pipil, Tuscarora, Yaqui, and Mandarin (22 sampled languages; Pipil colexifies 'straw for hatching' more specifically, and in Efik, 'straw' and a particular type of grass are colexified). Complex terms involving a term meaning 'dry' or 'withered,' as for instance in Yoruba *koríko gbígbẹ* 'grass dry,' are found also in Nivkh, San Mateo del Mar Huave, Fijian, Great Andamanese, Hawaiian, Samoan, and Tetun (Meyah has *mofombrá efej* 'weed dry'). Variants are Kolyma Yukaghir *jaqada:čəd-ulegə* 'horse-grass' and Cheyenne *mo'e'évôhkêha'e* /mo'e'é-hóhkêha'e/ 'grass-hat' (note that in Pawnee, *kaʔiihcuʔ*can refer to both 'straw' as well as a 'straw hat' or 'straw basket'). One of the relevant Fijian terms *kaunisilāmadū* consists of *kau* 'wood, stick' the possessive marker *ni*, *silā* 'maize-like plant, maize' and *madu* 'dry;' a similar term is also found in Yoruba: *igi agbado tabi bàbà* 'wood maize or guinea.corn,' and Carrier has a semianalyzable term where one constituent means 'wood' or 'stick.' Further, there are also other languages in which 'straw' is lexically associated with 'corn' or 'maize': Wayampi *awasi-i-ɛ* is analyzable as 'maize-leaf/stalk-PAST' (referring to 'straw of maize' specifically), Chayahuita has *shi'shi' sha'huëtë pochin ninin-so'* 'maize leather/bark like do/be-3SG.SUB,' and Khoekhoe features a semianalyzable term with a constituent meaning 'corn.' Greek, Lake Miwok, and Embera colexify 'straw' with 'chaff,' and Khoekhoe, Muna, and Basque with 'stalk' and/or 'cane' ('stalk of corn' specifically in Khoekhoe). Similarly, Dongolese Nubian has *hasénnkášš(ɪ)* 'refuse of reaped crop, straw,' containing *hásɛd* 'stumps of crops remaining after harvest, stubble' and *kášš(ɪ)* 'refuse, waist (of vegetation).' In Khalkha and Abipón, presumably by provenience contiguity, 'straw' and 'mat' are colexified. Aguaruna appears to colexify 'straw' with 'layer of grass' and 'lawn,' Cavineña colexifies 'pasture,' and Embera *phõãrã*, associated with another gender, also means 'pasture, grassland.' In Tetun, the association

is present due to colexification of 'grass' and 'pasture' by an analyzable term with the other constituent meaning 'dry.' Finally, Badaga, Sora, and Guaraní colexify 'straw' and 'fodder.'

Other associations include: the plural form of Buli *wuuk* 'grass, blade of grass, straw' means "bushland, grassland, bush," and Abzakh Adyghe *χə-pqʔe* is analyzable as 'mow-place' and also denotes a 'field after harvest.' Blackfoot *soi'stsipínnakssin* is derived from *soi'stsipínnaki* 'to harvest' and indeed also denotes a 'harvest.' Nez Perce *peqes* denotes, alongside 'straw,' the 'Bunchgrass' specifically as well as "wild wheat, wheat." Wappo *pá·haʔ* also means 'belt' (presumably due to phonological collapse of Span. *faja* 'strip, waistband' with *paja* 'straw'), and Wintu *sunus* also means 'nest, lair.' Another Wintu term, *łEh*, means inter alia 'bed' alongside 'spread, hay.' Central Yup'ik *cupun ~ cup'un* is analyzable as /cupe-n/ 'blow-instrument' and can also refer to 'embers' and a 'rifle.' Miskito *rais mina* is analyzable as 'rice husk' (*mina* in fact also denotes the 'foot' and by extension also the lower part of something), while *waha*, another Miskito term for 'straw,' colexifies 'leaf.' Ancash Quechua *achu* also means 'splinter,' Tsafiki *yaja* 'straw for houses' contains *ya* 'house, roof,' and Bwe Karen *ɓu lʊkhwa* contains *ɓu* "paddy, unhusked rice;" the language also has another semianalyzable term containing an element meaning "(on) the upper or outer surface of."

60. The Sun

Representation in Dabase: 99%
Motivated: 51.8%
Thereof Analyzable: 5.8% Thereof Colexifying: 46.0%
Thereof by Contiguity: 25.7% Thereof by Similarity: 10.9%
Recurrent associated meanings: day, moon, clock, sunshine/sunlight, time, hot/heat, hour, sail, calendar, noon, weather, god, eye, burn, sunray/sunbeam, year, fire

The most frequent association between the 'sun' and other meanings is that with 'day, daytime' by contiguity. Colexification is found in as many as 37 languages, namely Hausa, Mbum, Rendille, Buin, Gurindji, Kyaka, Mali, Ngaanyatjarra, Sko, Toaripi, Japanese, Sora, Kildin Saami, Cahuilla, Ineseño Chumash, Comanche, Itzaj, Kiliwa, Lake Miwok, Pawnee, Quileute, Xicotepec de Juárez Totonac (by the complex term *chi'chiní'* /chi'chi-ni'/ 'warm-AGT'), Yana, Yaqui, San Lucas Quiaviní Zapotec, Copainalá Zoque, Lengua, Miskito, Rama, Tehuelche, Wichí, Fijian, Great Andamanese, Hawaiian, Manange, White Hmong, and Sedang. Similarly, Kyaka and Tasmanian (Southeastern and perhaps Middle-Eastern varieties) colexify 'sun' with 'daylight.'

Frequently, 'sun' is also extended to convey other time-related concepts. Indeed, the very notion of 'time (of day)' is expressed by the same term as 'sun' in Ngambay (here also 'moment' is colexified), Buin, Burarra, Badaga, Cahuilla, Itzaj, and Tehuelche, and in Ngambay, Highland Chontal, and Cashinahua (here also colexifying 'year,' as in Wayampi, and 'brilliance, strength of sun'), the relevant terms also mean 'hour' (note that none of these languages also colexify 'day'!). Some languages, namely Burarra, Gurindji, Toaripi,

Upper Chehalis, Cheyenne, Comanche, Kashaya, Kiowa, Wappo, Wintu, Bora, Cubeo, Jarawara, Yanomámi, and Kapingamarangi, employ their word for 'sun' directly also to denote a 'clock' or 'watch,' and in Cheyenne and Wappo also 'calendar' (and there are many languages where 'clock' is expressed by a complex term involving 'sun,' see section 79). In Sko and Buli, 'noon' specifically is colexified with 'sun,' and Buli and Itzaj feature terms which can also convey the meaning 'weather' and/or 'climate' (Buli also colexifies 'season').

As an alternative to the colexification of 'sun' and 'day,' there are languages with analyzable terms for 'sun' on the basis of 'day,' most frequently in Southeast Asia and Oceania (Urban 2010). Sahu has *wangere ma la'o* 'day POSS eye,' Yay *taaŋ*^{1} *van*^{4} 'obstruct.view day,' and Fijian, alongside *siga*, which colexifies 'day' and 'sun,' also has the complex term *mata-ni-siga* 'eye/face-POSS-day/sun.' San Mateo del Mar Huave has *teat nüt* 'father day.' The association with 'eye' is also found in Malagasy (*masonàndro*, analyzable as /màso-n-àndro/ 'eye-GEN-day'), and a somewhat similar term is found in Vietnamese (*mặt trời* 'face sky'); note also Tetun *loro-matan* 'sun-eye' for 'disk of the sun' specifically. Furthermore, there is evidence for a diachronic association between 'day' and 'sun' in Basque (compare *eguzki* 'sun' and *egun* 'day'), and for a connection between 'eye' and 'sun' in Irish, where the present-day word for 'sun' is cognate with the inherited Indo-European word for 'eye' (Mallory and Adams 2006: 128). A Middle-Eastern Tasmanian term for 'sun' is recorded as *pŏganubrenā* which resembles *pŏga(na)* 'man' and *nūb(ĕ)rē(na)* 'eye.' However, the simplex terms are not attested for the same language or dialect but for different ones, and thus it is unclear whether the mentioned term can be analyzed in this way or whether a diachronic connection between 'eye' and 'sun' should be recognized. A different recurrent pattern is colexification of 'sun' with 'warm, warmth' or 'hot, heat,' or complex terms exhibiting this association (see Buck 1949: 54 for the possibility of this connection in Irish). Xicotepec de Juárez Totonac *chi'chiní'* was already mentioned. Similarly, Anggor has *hüfü-hamindɨ* 'hot-bone/very,' the Nunivak island dialect of Central Yup'ik has *puqla-neq* 'warmth/heat-thing.that.results.from' (attested presently only in the Nunivak Island dialect, but recorded in the 19th century also for other dialects), and Abipón has *m-pae-Ra* 'POSS.INDEF/3SG-hot-ABSTR.' This term colexifies 'sun' with 'heat (of the sun),' and this pattern is also found utilizing monomorphemic terms in Buli, Miskito, Rama, and Hawaiian (among other meanings in this language). Relatedly, Dadibi and Japanese colexify 'sun' and 'fire' (the Japanese terms are different in prosody though; Dadibi *sia* in addition denotes the sensation of heat from fire or the sun), and in Lavukaleve and Ancash Quechua, the relevant terms have a verbal usage, in which they assume the meaning 'to burn.' In addition, Piro has the semianalyzable term *tkatši* which appears to contain *tši* 'fire.' Finally, many languages of the Americas use the same term for both 'sun' and 'moon' (and concomitant extensions typical for 'moon,' such as that to 'month' in some languages). Sometimes languages have complex terms for 'moon' based on 'sun,' see section 38. This phenomenon is encountered among the languages of the sample in Blackfoot, Carrier, Upper Chehalis, Cheyenne, Chickasaw, Lakhota, Nez Perce, Nuuchahnulth, Quileute, Tuscarora (according to older sources incorporated into the consulted source), Wappo, Wintu, Bora, Cayapa (here colexifying also 'lowered, landed, fallen'), Cubeo, Hupda, and Macaguán (for further

associations due to the 'moon'-reading, see section 38). In some of these languages, while the semantics of the colexifying term clearly allows reference to both heavenly bodies, there are complex terms to disambiguate, for instance Lakhota *wí* 'sun, moon,' *ą́pa-wí* 'day-sun/moon,' *hąhépi-wí* 'night-sun/moon;' the complex term for 'sun' of this structure in Cubeo can also be used to refer to a diurnal animal.

In ten sampled languages, Buli (also by the variant term *wen-biri* 'sun-seed'), Dongolese Nubian, Baruya, Ngaanyatjarra, Rotokas, Waris, Basque, Lesser Antillean Creole French, Embera, and Bislama, 'sun' and 'sunshine' or 'sunlight' are colexified, and Dongolese Nubian and Nuuchahnulth colexify 'sun' with 'sunray, sunbeam.' Finally, Aguaruna *étsa* also is the name of a mythological hero and the god of hunters as well as the name of yellowish bird and fish species (note that Buli and Highland Chontal colexify 'sun' with 'god,' and that the same term in the former language is also a "religious concept denoting the 'alter ego' or 'personal god' of an individual"). Similarly, Bororo also uses its term for the 'sun,' *meri*, as a name for certain spirits. Hawaiian, Kapingamarangi, and Samoan colexify 'sun' with 'sail' (according to Elbert and Pukui 1986: 188, in the sense of 'sun' the relevant Hawaiian term *lā* goes back to Proto-Polynesian **laʻaa*, and in the sense of 'sail' to **laa*).

Other associations are: Buli *wen* colexifies 'sun' with 'sky' as well as 'up, upwards,' while Koyraboro Senni *woynow ~ weynow ~ woyna* might contain *woy* 'woman, female' (which would then in all likelihood be motivated by mythological complexes; note also *woynow* 'hemorrhoids'). Buin *rua* also means 'door' (accidentally, due to borrowing from English) as well as "Be reddish-yellow (the colour of the sun at dawn)," and Burarra *marnnga* is the only term in the sample which simultaneously denotes both 'sun' and 'star' (as well as a type of shellfish similar in appearance to a star and several types of sea urchin; the language also has an unrelated monomorphemic term for 'star'). Kwoma *ya* also means 'decoration' and 'money' among many other things, Nunggubuyu *aḻir* is also used to refer to the 'starfish,' and Yir Yoront *pung* is said to also mean 'payback pendant.' Individual variants of Badaga *oṭṭu ~ ottu ~ hottu* can also mean 'to adhere,' 'pod, vegetable waste,' and act as a suffix "indicating numerical frequency." Greek *ílios* also means 'sunflower,' Japanese *taiyō* also 'ocean,' Ket *ī* also 'name,' Cheyenne *ešeʹhe* also 'compass,' and Kiowa *pʜe* is also used to convey the meaning 'summer.' The literal meaning of Nez Perce *wiyetenéʻt* is 'the one that travels' (to have terms for celestial bodies derived from a verb meaning 'to travel' is common in the American Northwest). Nuuchahnulth *hupaɬ* also is used to refer to the 'thimbleberry,' and Wintu *tulcuheres*, which is at the same time the name of a mythological hero, might be, with additional information from a Wintu myth, analyzable as "the one who was beaten as a (potential) spouse," compare *tul* 'to beat a spouse.' Central Yup'ik (Norton Sound dialect) *macaq* also means 'to shine.' Kaingang *rã* also means 'near close,' 'below,' and 'to begin, be about to,' Jarawara *bahi* also denotes 'thunder' and 'lightning,' whereas the Toba term *i-coỹadelec na* is analyzable as '3-illuminate 3SG.' Wayampi *kwalaɨ* also means 'dry season,' Bislama *san* also rarely is used with the meaning 'son' (due to collapse of Engl. *sun* and *son*), Hani *naolma* might contain *ma*, a classifier for big things, and Rotuman *asa* also means 'reputation, honor' inter alia.

61. The Swamp

Representation: 71%
Motivated: 42.0%
Thereof Analyzable: 25.3% Therof Colexifying: 16.9%
Thereof by Contiguity: 12.3% Thereof by Similarity: 3.7%
Recurrent associated meanings: pool/pond/lake, mud, water, puddle, dirt, wet/moist, clay, reservoir, damp, earth, meadow/grass, lagoon, pus

Terms for 'swamp' ('fen,' 'marsh,' 'bog,' 'mire') are frequently associated with 'pond, pool,' or 'lake,' either because languages do not distinguish lexically between the two and boundaries are somewhat fluid, or because 'swamp' is expressed by a complex term on the basis of one of these meanings. Colexification is found in Buli (where the meaning 'lake, pond' is rare), Ngambay, Gurindji (where the relevant term is also the name of a specific lake and denotes "something in ceremony" additionally), Kwoma (also colexifying 'waterway, canal' and 'lagoon,' which latter meaning is also colexified in Yanomámi, while in Tetun *kolan* 'lagoon' is in addition glossed as 'saltwater swamp' in parentheses), Basque, Ket, Chickasaw, Lesser Antillean Creole French, Pawnee (by a semianalyzable term containing an element meaning 'to be a place'), Kaingang, Yanomámi (by the term *wawëwawë*, reduplicated from *wawë* 'wide, empty), and Hani (by a semianalyzable term containing an element meaning 'sea, ocean' and 'to soak in water'). Similarly, Sedang colexifies "stagnant pond" more specifically. In the case that analyzable terms are found, the second constituent can mean 'mud,' as in Baruya *ara'bunya* /araka-bunya/ 'mud-pond' and also in Kyaka (there is also a variant term where additional constituents are present), 'dirt' as in Hawaiian *ki'o lepo* 'pool dirt,' 'wet' as in Mandarin *zhao3-ze2* 'pond-watery.terrain/wet,' 'spoiled' as in Meyah *méren óbóha*, which is apparently analyzable as /méren obohá/ 'lake spoiled,' or 'earth' as in Khalkha *cøgerym γazar* 'pool/lake earth.' In addition, Ngaanyatjarra *yurungarri* consists of *yuru* 'lake' and *-ngarri*, glossed as 'those associated with,' and in Ket, there are several terms for 'swamp' containing one for 'lake' inflected for plurality (see § 4.5.2.1. on the quasi-derivational function of the plural morpheme in Ket). A term for 'swamp,' *yoḳa´ ~ ayoḳa´*, that is apparently formed by adding *ka* 'something' to *yohi ~ ayohi ~ ạyohi´ ~ hayo´ha ~ ayo ~ ayox* 'lake' is found in Biloxi. There are also several languages in which 'swamp' or 'marsh' are colexified with 'puddle,' namely Muna, Basque, Khalkha ("mud puddle" more specifically), San Mateo del Mar Huave (the relevant term *ndorrop yow* is analyzable as 'hole water'), the Cuisnahuat dialect of Pipil (by the analyzable term *ta:l-a:-pu:ni* 'ground-water-be.born,' which also means 'spring of water,' compare Upper Chehalis *mó·lx*^{w} 'swamp' and *mó·l-* 'spring'), Arabela, and Bora. A Chickasaw term for 'swamp' has the idiolectal meaning 'puddle.' Further, in Miskito, *piahka tara* is analyzable as 'puddle big,' and in Hawaiian *ki'o lepo* as 'pool/puddle dirt.'

Several of the associations arising in complex terms with 'lake' also occur in other configurations. Buli, Hausa, Yoruba, Kosarek Yale, Basque, Khalkha, Arabela, Hupda, Ancash Quechua, Hawaiian, and Lenakel directly colexify 'swamp' with 'mud' ("[m]ud which is so deep as to be practically impassable" in Hausa and 'black mud' specifically in Hupda), Baruya has *ara'darya* /araka-'darya/ 'mud-area,' Kyaka *ipwua manduwua* 'water

mud' (this term also denotes 'watery mud' and 'slush'), and San Mateo del Mar *n-ajndor iüt* /na-ajndor iüt/ 'AGT-be.muddy-ground.' Further, Cubeo *yao-bu* is derived from *yaa* 'mud, adobe' by means of the classifier *-bu* for hard and/or round objects, Kaingang has *óré ki goj* 'mud in water/river,' and Fijian *vanua oruoru* 'land muddy' (Fijian also has *lōlobo*, reduplicated from *lobo* 'to stick in the mud, to penetrate something'). Furthermore, Guaraní *tuju rapo* and *tuju rupa* contain *tuju* 'mud, abode.' The relevant Khalkha term also colexifies 'clay;' so do the Buli and Ancash Quechua terms. Likewise, terms for 'swamp' in which the meaning 'earth' figures are not only attested in conjunction with 'lake': Efik has *memmem isöñ* 'soft earth/ground,' and in Embera, *egoró-susúa* contains *egoró* 'ground, earth' and *susúa* 'watering place;' a semianalyzable term including a constituent with the meaning 'earth' is also found in Bora. Alongside the association with 'dirt' by the overt term in Hawaiian, Buli, Kosarek Yale, and Rama colexify 'swamp' with 'dirt.' There are also many languages with complex terms for 'swamp' where one of the constituents means 'water,' some of which were already mentioned above. Kanuri has *njì-bòné* 'water-lie.down' and Kyaka *ipwua manduwua* 'water mud,' *ipwua wara pete* 'water watery.mud pool' and *ipwua/ipya inju malu singi dokona* 'water mud much usual.location there.' Ngaanyatjarra *kapingarri* (meaning "area where water runs, swampy area") consists of *kapi* 'water' and *-ngarri* 'those associated with,' Abzakh Adyghe *psə-čʲe=psə-λe* is analyzable as 'water-source=water-DERIV,' Ket *ulteɣin* as /ul-te-in/ 'water-lake-PL,' Sora *duŋdəm'da:'lo:n* as /duŋ-dəm-dˀa:-'lo:-n/ 'get.out.of-REFLX-water-ground-N.SFX,' Haida *xawtl'adaanggaa* 'be swampy' contains *xaw* 'liquid' and *daanggaa* meaning 'have been discarded' inter alia, Kashaya *qʰaṭʰuʔul* is analyzable as /ahqʰa-ṭʰuʔul/ 'water-old' (and also denotes "stagnant water" generally), Nez Perce *kusíˑn wéˑtes* as /kuˑs-íˑn wéˑtes/ 'water/dew-with land,' Pipil (Cuisnahuat dialect) *ta:l-a:-pu:ni* as 'ground-water-be.born,' Wintu *meˑm λat-i* as 'water/river wet/damp/moist/soak/drench-??,' Yaqui *ba'a jeelo* as 'water near,' and Wichí *inot-w'et* as 'water-place;' semianalyzable terms where one of the constituents can be identified to refer to 'water' are found in Blackfoot, Chickasaw, Comanche, Kashaya, Yuki, Cavineña, and Piro.

Furthermore, 'swamp' is associated with 'meadow' and/or 'grass' by colexification in Badaga and Nez Perce (in Badaga also with 'water channel' and "upper part of a ridge;" somewhat similar is White Hmong *hav iav* 'valley grassy' meaning 'wet lowland'). Lake Miwok colexifies 'to be swampy' with 'to be damp,' and in two sampled languages, Chukchi and Wintu, there are analyzable terms for 'swamp' in which the meaning 'damp' figures: *ilə-lqen* 'damp-on.top' and *mem λat-i* 'water/river wet/damp/moist/soak/drench-??' respectively (compare also Japanese *shitchi*, analyzable as /shitsu-chi/ 'moist-place,' Mandarin *zhao3-ze2* 'pond-watery.terrain/wet,' and colexification of 'swampy' and 'wet' inter alia in Hawaiian). In two sampled languages, Central Yup'ik and Rama, a metaphorical association with 'pus' is found: in Central Yup'ik by the analyzable term *imarrluk*, consisting of *imaq* 'contents, bullet, pus, ocean' and the postbase (see § 4.4.2.) *-rrluk* 'one that has departed from its natural state,' and in Rama by colexification (also with 'dirt'). Finally, Badaga, Basque, and Khalkha, colexify 'swamp' with 'reservoir.'

Other associations include: Buli *biung* is primarily the name for a "watering hole that dries up quickly," and *viak* also means 'valley.' Hausa *damba* is also used with the

meaning "a trap in speech" inter alia, Ngambay *bbar* also means 'sound, noise,' and Swahili *kinamasi* also 'slime.' Berik colexifies 'swamp' with 'delta,' and Buin *kapunuka* is also an "insulting term for a woman." Burarra *mugurrmulmul* is analyzable as 'CLASS.DOMESTIC-paperbark.tree' and denotes a "swamp area" and "in a stand of small paperbark trees," and the Kwoma term *biira-biira she* is analyzable as 'open-RED faeces/ash.' Yir Yoront *walq* also means "hollow place, concave surface" and "main (deep) waterhole." Badaga *oḷe ~ hoḷe* also means 'river' and Basque *istil* also 'waterhole.' Khalkha *sibar namuγ* also denotes 'plaster, stucco,' and *taca* also 'broom' and the 'Golden Chain' (*Cytisus laburnum*). Chickasaw *okpachalhlhi'* also means 'mudhole.' Welsh *siglen* is derived from *sigl* "a shaking, oscillation." Itzaj *k'al* also means "closing, closed in place," and the Tuscarora root *-hnaw-* occurring in *uhnà·weh* 'morass, swamp' also yields the meaning 'current of water,' in particular when occurring incorporated. Central Yup'ik (Yukon dialect) *puglerneq* is analyzable as /puge-ler-neq/ 'come.to.surface-suddenly-thing.that.results,' and for the Bristol Bay dialect term *angayaq*, compare *angala-* 'to flutter, wobble.' Tsafiki *mudú* appears to be analyzable as /mu-du/ 'achiote-mountain,' and Tehuelche colexifies 'swamp' with 'skullcap' or 'wit' (original gloss is 'mollera'). Hawaiian *pohō* also means 'sunken, sinking' inter alia, *naele* also includes 'rock, crevice' in its denotational range, and *nenelu* also means 'flabby fat' or 'soft,' again among other meanings. Malagasy *hòraka* also denotes a "wet rice field," and Manange *1tsʰo* also a 'rope.' Samoan *taufusi* is also used to refer to a "patch of ground irrigated for the purpose of growing taro," and *pala* also means '(be) rotten' and 'to decay, perish' inter alia when used as a verb. Sedang *lông* also means 'to sing a lullaby, put a child to sleep, coax.'

62. The Tail

Representation: 95%
Motivated: 21.5%
Thereof Analyzable: 3.8% Thereof Colexifying: 17.7%
Thereof by Contiguity: 5.6% Thereof by Similarity: 14.9%
Recurrent associated meanings: end, penis, tailbone, buttocks, back, train of dress, tail of cloth, tail of coin, tailfeathers

While there are some morphologically complex terms for 'tail' in the languages of the sample, recurrent associations are almost exclusively realized by colexification. Most commonly, languages use their word for 'tail' also for 'end' in general (paralleled in the diachrony of Indo-European in Irish, Buck 1949: 209). This is found in twelve sampled languages, namely Buli (where the relevant term also is the name for the 'fly-whisk' inter alia), Yoruba (also colexifying 'completion, conclusion'), Kwoma (colexifying "lower end" or "downhill side," as well as "the lower of two entities" more specifically), Kosarek Yale, Abzakh Adyghe, Basque, Khalkha, Lesser Antillean Creole French, Maxakalí, Rama, Kapingamarangi, and Samoan, in which latter the relevant term also means 'to finish' as a verb inter alia. Similarly, Sedang colexifies 'tail' with "end of cord, loincloth," and Efik *ntañ ëtak* is analyzable as 'feather end.of.body' (this term denotes the tail of birds specifically, for tails of quadrupeds and fish, there is a separate monomorphemic term. It should

be noted that in a number of sampled languages, different unrelated terms are used for the tails of specific types of animals). Moreover, Samoan *siʻusiʻu* is reduplicated, the base being *siʻu* 'tip, extremity.' More specific associations go in the same direction: Rendille, Basque, and Kolyma Yukaghir colexify 'tail' with 'buttocks' (furthermore, Ngaanyatjarra colexifies 'tail of insects' specifically with 'bottom, buttocks,' compare also Pipil (Cuisnahuat dialect) *-kwitapil*, which is derived from *-kwita* 'excrement' by means of the (frozen) diminutive suffix *-pil*). Moreover, in Khoekhoe, the same root yields the meanings 'tail' and 'buttock,' with different nominal designants distinguishing the two. In Rendille, Abzakh Adyghe, and Wichí, terms for 'tail' also may refer to the 'back' or 'backside' of something, and Greek and Hawaiian use the same term for 'tail' and 'train of dress.' Similarly, Hausa and Abzakh Adyghe colexify 'tail' with 'tail of cloth' (as well as "a small quantity of sour milk given free to a purchaser of butter" in the former language). The relevant Hausa term is also extended to 'penis,' a pattern also attested in Koyraboro Senni, Lavukaleve, and Cashinahua, and perhaps in Anggor, where this meaning is marked with a question mark in the consulted source (compare the cognacy of New Persian *dum* 'tail' with Old High German *zumpfo* 'penis' noted by Buck 1949: 210; this pattern seems to be widespread in languages of Europe generally, for instance, Latin *penis* originally had reference to the tail of animals). There also is a term for 'tail' where the identifiable constituent is that for 'penis' in Berik. Rotuman *reu* also denotes the 'tail of a coin' (and also the 'foot of a bed'), and a term for 'tail' in Hausa is similarly used to refer to specific motifs on the back side of coins. Relevant terms in Cheyenne, Haida, Macaguán, and Yanomámi simultaneously denote 'tail' and 'tailbone, coccyx' (and the Yanomámi term also 'stinger'), while Nez Perce and Wayampi colexify 'tail' with 'tailfeathers.'

Other unsystematic associations include: Buli *jiuk* also denotes a "bird's trap made of grass," Dongolese Nubian colexifies 'tail' with 'to wash' and 'to send,' while Yoruba *apa ẹ̀hin ohunkóhun* is analyzable as 'part/arm back whatever.' Baruya *suya* also means 'urine,' and for Kaluli *waf* 'tail,' compare *wafe* 'worm.' As a verb, Muna *punda* also means 'to jump,' and *lensi* also 'to untie, undo, loosen,' while relevant Tasmanian terms in all varieties except the Northern one appear to denote any excrescence of the body: 'wart,' 'scar,' 'wrinkle,' and 'tail.' Sko *pú* is also the name of a furry mammal, and Yir Yoront colexifies 'tail' with "tail-end of spear." The Abzakh Adyghe term *qᵒʔeps* is analyzable as /qᵒʔ(e)-ps(e)/ 'branch/pointed.object-string' and colexifies 'strap' and 'handle,' and another term in the same language can also refer to an 'egg,' a 'seed,' and other things. Badaga *ba:lu* also is used with reference to a 'meteor,' "slender means," and "anything meager." Basque *buztan*, colexifying 'end' and 'butt,' also means 'shoot, sprout,' and *isats* also 'broom' and 'consequence, upshot.' Chukchi *ŋojŋən* also means 'rump,' whereas Greek *ourá* also means 'queue.' Welsh *llosgwrn* appears to contain *llosg* 'burning,' Cheyenne *hehévá'xe* is also the name for a piece of meat, the 'oxtail.' San Mateo del Mar Huave *wiül* also means 'vixen,' and *coy* also denotes the 'coney' as well as a picture or drawing of any animal and 'rheumatism.' The Nuuchahnulth term *načа* also means 'fluke,' Santiago Mexquititlan Otomí *ts'u̲* also 'nipple, teat,' and Central Yup'ik (Norton Sound dialect) *pamyuk ~ pamsuk* also 'tail of canoe' and 'chorus of song.' Carib *-antɨkɨlɨ* is analyzable as /-antï-(e)kïlï/ 'behind-spin' (compare the origin of German *schwanz* from *sweifen* 'turn around,' Buck 1949: 210). Guara-

ní *tuguái* may also refer to an 'appendix' and metaphorically to 'family, offspring,' Ancash Quechua *chupa* also is used to refer to a "person who is always behind someone else," and Tehuelche *t'e:r ~ t'er* also means 'bristle.' Fijian *bui* is also used for 'grandmother' and 'old gossip,' Hawaiian *huelo* also has the figurative meaning 'inferior,' and Bwe Karen *ká=mè* is analyzable as 'hind.part=knot/joint.' Lenakel *nɨpɨk-* also denotes the 'tail of a stingray' and 'stern of canoe or boat,' and Manange *1mẽ* also means 'fin.' Sedang *tíng* also means 'to sacrifice to the spirits,' *xô´i* has similar other meanings, Tetun *ikun* is also used as a term for the youngest of somebody's children, Vietnamese *đuổi* also means 'to pursue,' and Yay colexifies 'tail' with 'head of rice.'

63. The Thorn

Representation: 72%
Motivated: 42.2%
Thereof Analyzable: 7.8% Thereof Colexifying: 34.4%
Thereof by Contiguity: 8.0% Thereof by Similarity: 18.4%
Recurrent associated meanings: needle/awl, splinter, quill, point, sharp, stinger, thorny plant, sticker, tooth, barb, beak, injection, penetrate/pierce, nail

Most frequent cross-linguistically is colexification of 'needle' or 'awl' and 'thorn' (or 'spine,' 'prickle'), which may be either due to provenience contiguity or to perceptual similarity. This pattern is attested in Anggor, Baruya (in these languages colexifying also 'injection;' 'give an injection, inject' is also colexified in Samoan), Buin, Khalkha, Abipón, Bora, Chayahuita, and Yanómami (in the latter language, the relevant term also means 'pin;' Tehuelche *xolnwe ~ xoln*, furthermore, contains *xol ~ xo:l ~ ʔexol* 'to sew' and colexifies 'nail' in addition, as does Bislama).

Otherwise, colexification with other sharp pointed objects is frequent. Kyaka, Muna, Nunggubuyu, Kolyma Yukaghir, Cheyenne, Wintu, and Tetun colexify 'thorn' with 'splinter' (Kyaka also with 'excrescence' generally), Kaluli, Kosarek Yale, Highland Chontal, Bororo, and Samoan with 'stinger,' and Kyaka, Kosarek Yale, and Lengua with 'tooth' (Kyaka also with "biting, erosive" and 'food' inter alia). Note also that Maxakalí *xãpxox* presumably contains *xap* 'stone, bead, seed' and *xox* 'tooth, sharp fragment.' Hausa, Ngambay, Haida, Lesser Antillean Creole French, Central Yup'ik, and Yanomámi colexify 'thorn' with 'quill,' Kyaka, Bororo, and Lengua with 'beak,' Cahuilla, Lake Miwok, Pawnee, and Wintu with 'sticker,' while the relevant terms in Kosarek Yale, Bororo, Guaraní, and Miskito are extended to refer to the 'point' of an object more generally, and Ancash Quechua also to 'anything pointed.' In addition, Yoruba colexifies 'thorn' with 'bone,' while Cayapa colexifies 'thorn' with 'slim and acute bone' more specifically.

Efik *ñku'kïm* (containing *kïm* 'to pierce') is also used to refer to a 'spine on a shell' (alongside 'patchwork'), Basque *arantza* also means 'spine of an animal,' and Hausa *k'aya* also has the meaning 'fishbone.' Koyraboro Senni *karji,* Hawaiian *kukū* as well as Samoan *tala* also mean 'barb' (*kukū* also means 'burr' as well as 'to hurt by a thorn' and 'to hit' inter alia, and *tala* also 'prong' and 'spur'), and Ngambay's *hay* also is the term for a 'particular kind of straw' (alongside 'paddle'). Ngaanyatjarra colexifies 'spike' (compare Samoan *tuitui*

'thorn' with *tui* 'to stab, jab,' 'spike'). Kyaka and Guaraní colexify 'horn,' Rotuman *kō* colexifies 'fork' (a pattern shared with Samoan, *kō* also means 'to stab, pierce,' and the relevant Samoan term also has other meanings), and Bislama *nil* colexifies 'cock's spur.' For Lake Miwok *kíili*, which means "to hook with the horns" as a verb, compare *kílli* 'horn, antler;' the Toba term *le* also means 'tooth of a comb' and 'point of a lance.' The semantic feature of 'sharpness' which in all likelihood underlies these patterns of colexification is made explicit in other languages by complex terms. Upper Chehalis *łač=áχn* is analyzable as 'sharp=edge' (and indeed may also refer to a 'sharp edge'), Chickasaw *naa-haloppa-'* as 'something-be.sharp-NMLZ' (note also that there is a semianalyzable term involving a constituent meaning 'thing' in Bwe Karen), and Hawaiian *ʻoiʻoi* is repuplicated from *ʻoi* 'sharp' (this term may also be used to refer to a 'superior person' and assumes the meaning "to protrude, stick or jut out" in verbal usage). A semianalyzable term containing a lexical element with the meaning 'sharp' is also found in Pawnee, and Kyaka and One directly colexify these meanings. Relatedly, Rotokas and Xicotepec de Juárez have terms for 'thorn' derived from verbs meaning 'to penetrate' and 'to pierce' respectively; direct colexification of these meanings is found in Rotuman, and a semianalyzable term of this type is found in Efik. Another recurrent pattern is colexification of 'thorn' with a thorny plant, either with a particular one or generically. Thus Buli *mung* also denotes a thorn tree of the *Acacia* genus (alongside a kind of beetle), Basque colexifies 'thorn' with 'hawthorn,' for Kiowa *sę̨ígā't* compare *sę̨ígā* 'cactus, peyote,' Tuscarora and Miskito colexify 'thorn' with 'thistle,' Tehuelche *čo:rč ~ čo:rč' ~ č'o:rč' ~ čorč* also denotes the 'Calafate shrub' (*Berberis microphylla*), and Pawnee, Wintu, Bora, and Wayampi colexify 'thorn' with 'thorny plant, thorny bush' in general.

Other associations include: Muna *kiri*, as a verb, also means "to insert a thorn into something" as well as 'to scrape off,' and Ngaanyatjarra *ngunyarrma* also means 'rasp.' Kosarek Yale *si* also means 'name,' and *alok*, another Kosarek Yale term, also 'for earth to slide off' and 'make hollow underneath something.' Kolyma Yukaghir *nono* also means 'handle,' Cahuilla *čuŋal* also 'jumping cactus,' and Bororo *oto* also 'peak, front.' Cavineña *acui-ja* is analyzable as 'tree-GEN,' and Miskito colexifies 'thorn' with 'twinge.' A literal translation of the Rama term *kú up* (/ku up/) would be 'bird's.wing eye,' and Tsafiki *po* is also used for trees of the Guadua genus. Fijian *voto* also means 'root of a body hair' and 'a hundred voivoi leaves.' Malagasy *tsìlo* also denotes the 'needles' of the pine tree, and Tetun *aitarak* also means "prickly, rough."

64. The Thunder

Representation: 91%
Motivated: 41.5%
Thereof Analyzable: 17.9% Thereof Colexifying: 24.5%
Thereof by Contiguity: 12% Thereof by Similarity: 28.1%
Recurrent associated meanings: lightning, god/spirit, sky, roar, noise/sound, (thunder)bird, cry/wail, storm, cloud, gun, rain, electricity, voice

21 sampled languages, Buli, Ngambay, Rendille, Yoruba, Kwoma, Abzakh Adyghe, Japanese, Itzaj, Xicotepec de Juárez Totonac, Yana, Yaqui, Central Yup'ik, Abipón (by the analyzable term *kahag-Ran-Ra* 'to.lighten-CAUS-ABSTR'), Arabela, Aymara, Hupda, Jarawara, Tehuelche, Bislama, Bwe Karen, Takia, and perhaps Middle-Eastern Tasmanian, directly colexify 'thunder' with 'lightning, thunderbolt' (compare the semantic shift from the latter to the former in Lithuanian, Buck 1949: 58). However, in contrast to the association with 'lightning,' complex terms for 'thunder' on the basis of 'lightning' are much rarer. San Mateo del Mar Huave *ajüy teat monteoc* is analyzable as 'walk father thunderbolt,' and Lavukaleve and Rama have semianalyzable terms for 'thunder' containing the respective term for 'lightning.'

For 'thunder' specifically, a general recurrent pattern is seen in terms derived from terms denoting some kind of loud noise (see Buck 1949: 57 for details on similar evidence from Indo-European languages). Terms in many languages contain a verb meaning 'to roar.' Mbum has *bálà-mbàm* 'roar-rain' and *fómà-mbàm* 'scolding-rain,' Guaraní has *arakororõ* 'sky-growling,' and Manange *3mo putul njut-si*, involving the constituents *3mo* 'sky,' *putul* 'dragon,' and *njut* 'roar.' Khoekhoe, Buin, Nunggubuyu, and Sora directly colexify 'thunder' with 'roar,' 'growl' and/or 'boom' (Sora also 'to shake'), and a semianalyzable term containing an element with that meaning is also found in Great Andamanese. Similar terms, in which, rather than 'growl,' more general terms for 'noise' or 'sound' figure as constituents are Katcha *(thimpidɔ) kafara* '(sky/rain) make.noise/cry,' Meyah *mocgój ogúgur* 'cloud noise,' Japanese *kami-nar-i* 'god-sound-NR,' and Tetun *rai-tarutu* 'earth-noise.' In Santiago Mexquititlan Otomí, Copainalá Zoque, and Bororo, the meanings are colexified, and semianalyzable terms with a constituent meaning 'noise' are featured in Copainalá Zoque and Abipón. Another class of motivated terms relating to some sort of noise is that comparing the sound of 'thunder' to 'wailing' or 'crying.' Katcha has *thafara ma thimpidɔ* 'wailing GEN sky/rain,' and Comanche *tomo-yaketʉ* contains *tomo* 'cloud, sky' and *yaketʉ* 'cry.' The Kiliwa term *?+kwiy=h+mii* has the same structure: it is analyzable as 'DN+cloud=3+cry,' White Hmong has *xob quaj* 'Xob cry' and *xob nroo* 'Xob moan,' and the meanings are colexified in Yir Yoront. In Carrier, the word for 'thunder,' *tl̓ît̓ni tetni*, contains *tl̓ît̓ni*, the name for a "gigantic bird" in Carrier mythology and *tetni* 'cry' (it is likely that *tl̓ît̓ni* is the Carrier incarnation of the 'thunderbird'). Similarly, Lakhota has *wakíyaŋhothúŋpi*, literally 'thunderbirds call.' Colexification of 'thunder' with a (mythological thunder-)bird is also attested in Waris, Upper Chehalis, Haida, and Miskito. Kiliwa also has an alternative term: *ha?=kw-?-ny+mar=kwiy* 'voice=WH-DN-POSS+image=cloud.' Similarly, in Pawnee the 'thunder' is called *wakuhtakaahak*; this term is analyzable as /wakur-tahaahak/ 'voice-drop.down. Moreover, Bakueri *ngálá lo̱wa* is analyzable 'gun god/sky' and, paralleling this association, Lavukaleve and Rotokas terms colexify 'gun' and 'thunder' (compare the extension to 'cannon' in Romanian, Buck 1949: 58). In Efik *erituak' en'yön* is analyzable as /erituak' en'yöñ/ 'knocking/beating.of.drum sky.' Still further, Hausa *aradu* also dialectally denotes "the wedge used in splitting palm wood," and *tsawa* also is used to refer to a 'loud rebuke' or "the cracking of newly-burned pots for no apparent reason" inter alia, while *cida* is also the name of a spirit, again inter alia. In Buin, *kururu* is also used for a "rumbling noise" in general and is also the name for a "large wooden trumpet" and a

"large buzzing fly," Bezhta has *hasa qäƛeyoɬi* 'sky:ERG shout:MASDAR,' and the Basque term *dunbots* also may refer to a 'din' or 'clamor.' Cheyenne *ma'heo'o énéstoohe* contains elements meaning 'god' and 'call.' Ineseño Chumash *'ašaqšk'apš* revolves around the verb *šaqšk'ap* 'to clap.' The final element is unkown. The Wichí term *pelhay y'iplhi* contains the lexical elements *pelhay* 'storm' and *y'ip* 'sing' (note that the same term in Embera yields the meanings 'thunder' and 'storm, tempest' when associated with different genders, and that 'storm,' 'rainstorm,' or 'thunderstorm' is colexified with 'thunder' in Ngambay, Rotokas, and Carib). A figurative Hawaiian term for 'thunder' is *ʻuʻina pōhaku a Kāne*, literally 'crackling rocks of Kane' (Kane being the principal Hawaiian god), with the element *pōhaku* also meaning 'thunder' by itself. Underlying this denomination, as well as *pohā-kaʻa* 'crack-turn' (though note for the first element that *pōhaku* has a short form *pōhā*) is the belief that thunder was caused by the gods rolling around rocks in the sky. As the discussion so far shows, in complex terms, 'cloud,' 'sky' or 'rain,' as meanings contiguous to 'thunder,' frequently figure in complex terms for that meaning. Alongside the terms in Meyah, Comanche, and Kiliwa, One, too, has a term involving cloud: *yemi piyale* 'cloud break' (and note that Badaga *iḍi* can also mean 'to break' inter alia). Alongside Mbum and Katcha, 'rain' also figures in Yaqui *yuku jimaa-ri* 'rain throw-RES,' Buli *ngmoruk* 'rain' rarely assumes the meaning 'lightning, thunder,' and Hawaiian colexifies 'thunder' with "raindrops, fine rain, to rain gently" inter alia, while Noni has a semianalyzable term involving 'rain,' and Aguaruna *ipamát* also denotes "to rain as a sign of death or an imminent attack" as well as 'to reveal a message in dreams.' Alongside Efik, Katcha, Bezhta, Guaraní, and Manange, 'sky' is also the meaning of one of the constituents of Hani *aoq-jiq* 'sky-sift.' An association with (a) god or a spirit is, next to Bakueri, and Cheyenne (where 'battery' is in addition colexified), also found in Japanese (*ika-zu-chi* 'horrible-GEN-spirit;' this term is archaic and the accuracy of the morphological analysis is questionable) and Yay (*pya*^{3} *ray*^{4} 'spirit cry'). Further, in Nez Perce, where *hinmé·t* is also the name of the "spirit of a cloud that makes noise" (alongside the reading as 'electric storm') as well as in Yana, Biloxi, Cheyenne, Tuscarora, Wayampi, and Yanomámi, there are associations with some kind of god or spirit by colexification. In Chayahuita, the word for 'thunder' is also the name of the person who castigates the evil after their death.

Central Yup'ik and Hupda colexify 'thunder' with 'electricity,' which is in all likelihood a sideffect of the fact that they also colexify 'lightning' using the same term; the association is much more frequent for 'lightning,' compare section 35.

Other associations include: for Anggor *burɨhoai*, compare *burɨ* 'many things existing upright.' *Nɨmamɨndohoafɨ* ~ *nɨmamɨndɨhoafɨ* contains *hoafɨ* 'to talk;' a literal translation provided by lexicographers is "above possessive taɨk [sic!]." Kyaka *yungala* also means 'praying mantis,' and Muna *tondu* also 'to sink, drown.' Rao *gramvuvre* appears to contain *gra* 'sun,' and Rotokas *varake-oto*, meaning 'thunder' as well as 'thunderstorm,' seems to be analyzable as 'very.high-to.punch.' Sentani *ku* also means 'bracelet,' and Blackfoot *ksiistsikomm* contains *ksiistsikó* 'day.' For Pawnee *kirir*, compare *kirir (uur...)* 'to shake, tremble.' Wintu *t*^{h}*um* is also used with the meaning 'coo.' Copainalá Zoque colexifies 'light thunder' with 'spark,' and Jarawara *bahi* also denotes the 'sun' (and, departing from there, 'clock, watch,' see Dixon 2004: 71), whereas Rama *dama yatangi* contains *dama* 'grandfa-

ther.' Fijian *kuru* also means 'to pursue' and 'to jostle,' Hawaiian *hekili* is metaphorically extended to also mean 'passion, rage' and also denotes a yam species. Samoan *fāititili* contains *tili* 'to tremble.' *Ta'a-lili* is fully analyzable: 'be.loose-tremble.'

65. The Tree

Representation: 97%
Motivated: 70.3%
Thereof Analyzable: 5.4% Thereof Colexifying: 65.3%
Thereof by Contiguity: 50.8% Thereof by Similarity: 0.9%
Recurrent associated meanings: wood, stick, plant, trunk/log/pole, shrub/bush, forest, branch, wooden artifact, thing, bone, splinter, canoe, gun/rifle

By far the most frequent lexico-semantic association is that with 'wood' (see also Buck 1949: 48 for this pattern in Indo-European). It is realized most frequently by colexification, in 83 languages of the sample, that is, in a little less than sixty per cent of sampled languages. This figure draws close to that arrived at in an earlier survey by Witkowski et al. (1981), who report that the pattern is found in two thirds of their sampled languages. In the present sample, colexification of 'tree' and 'wood' is found in Efik, Hausa, Khoekhoe, Koyraboro Senni, Ngambay, Swahili, Yoruba (also colexifying 'fuel'), Berik, Buin, Burarra, Dadibi, Gurindji, Kwoma (where the relevant term may also refer to a "wooden beater," "wood-carving," and "slit-drum, hollow log drum"), Kyaka (the term also means "human, earthly, mortal" and 'below, lower'), Mali, Ngaanyatjarra, Nunggubuyu, Meyah, Tasmanian (all varieties except the Northern one, for which data are lacking), Toaripi, Sahu, Sko, Waris, Kosarek Yale, Badaga, Japanese, Ket, Khalkha, Nivkh (colexifying 'firewood' and 'wooden' more specifically), Kildin Saami, Welsh, Kolyma Yukaghir, Biloxi, Cahuilla, Carrier, Upper Chehalis, Chickasaw, Highland Chontal, Ineseño Chumash, Comanche, San Mateo del Mar Huave, Itzaj, Kiowa, Lakhota, Nez Perce, Santiago Mexquititlan Otomí, Pawnee, Pipil, Xicotepec de Juárez Totonac, Wappo, Yana, San Lucas Quiaviní Zapotec, Copainalá Zoque, Aguaruna, Bororo (by the term *iguru*, for which compare *igu* 'rope' and *ru* 'fire'), Cashinahua, Cavineña, Cayapa, Embera (where the meanings are associated with different genders), Guaraní (where the relevant term can also refer to a 'plank of wood'), Huambisa, Hupda, Jarawara, Kaingang, Macagúan, Miskito, Piro, Tsafiki, Wayampi, Wichí, Yanomámi, Fijian, Hawaiian, Bwe Karen, Lenakel, Malagasy, Mandarin, White Hmong, Rotuman, Samoan, Takia, Tetun, Yay, and Bislama. In addition, there are a few sampled languages in which 'tree' is expressed by a morphologically complex term involving a constituent meaning 'wood.' Lesser Antillean Creole French has *pié-bwa* 'stem/foot-wood' and Manange *2ʃiŋ-3tuŋ* 'wood-copse/trunk.' Witkowski et al. (1981) propose that complex terms for 'tree' involving 'wood' can be seen as a quasi-evolutionary development, in which 'wood' is the more "salient" referent in small-scale societies which is first extended to 'tree' by colexification and then the 'tree'-reading is singled out by complex terms as societal complexity increases. In the light of this hypothesis, it is interesting to note that a Creole language in the present sample features such a complex term, which would entail that, if their general scenario is correct and also applicable to Lesser Antillean Creole

French, the development must have occurred in a very short time span, given that the lexifier language French does not colexify 'tree' and 'wood' (another possibility would be that Lesser Antillean Creole French acquired this pattern through relexification). In addition, in Upper Chehalis, *ƛ́íšaƛ́š* 'a clump of trees, woods' is reduplicated from the root *ƛ́íš-* 'wood,' in Cahuilla, *kélawat* 'tree, wood' is derived from the verb *-kélaw-* 'to gather wood,' and in Cubeo, 'tree' is *jocʉ-cʉ*, consisting of *jocʉ* 'wood' and the classifier for tree-like objects *-cʉ*.

Colexification of 'tree' with parts of trees also occurs. Mirroring the complex terms in Lesser Antillean Creole French and Manange, 'tree' and 'trunk,' 'log,' or 'pole' are colexified in 19 languages, namely Mbum, Kwoma, Ngaanyatjarra, Badaga, Chukchi, Ket, San Mateo del Mar Huave, Nivkh, Kiowa, Nez Perce, Oneida, Xicotepec de Juárez Totonac, San Lucas Quiaviní Zapotec, Aguaruna, Jarawara, Yanomámi, Hawaiian, and Kapingamarangi, 'tree' and 'stick' are in 34 languages (Efik, Khoekhoe, Ngambay, Noni, Burarra, Gurindji, Kwoma, Lavukaleve, Ngaanyatjarra, where "magic stick" is also colexified, Nunggubuyu, Yir Yoront, Chukchi, Ket, Khalkha, Kolyma Yukaghir, Biloxi, Blackfoot, Carrier, Upper Chehalis, Chickasaw, Comanche, Kiowa, Wappo, Yana, San Lucas Quiaviní Zapotec, Guaraní, Jarawara, Lengua, Maxakalí, Rama, Fijian, Hawaiian, Kapingamarangi, and Sedang), and 'tree' and 'branch' in Ngambay, Buin (colexifying 'small tree' and a particular tree species more precisely), Welsh, Blackfoot, and Kiliwa. Ngaanyatjarra and Hawaiian colexify 'tree' with 'splinter,' and Abipón is unique in colexifying 'tree' and 'bark' (as well as 'rose bush').

In Khoekhoe, Ngambay, Rendille, Ngaanyatjarra, Yir Yoront, Sora, San Mateo del Mar Huave, Kashaya, Kiowa, Nuuchahnulth, Quileute, Yaqui, Ancash and Imbabura Quechua, Fijian, Hawaiian, Rotuman, Samoan, Takia, Tetun, Yay (here the relevant term also means 'older brother'), and Bislama, 'tree'-terms are also used for 'plant' in general (though the Rendille term explicitly excludes 'grass' from its denotational range); similarly, Koyraboro Senni, Rendille, Kyaka, Nunggubuyu, Toaripi, Badaga, Lake Miwok, Nuuchahnulth, and Kapingamarangi also use terms for 'tree' for 'shrub, bush.'

There are also six languages, namely Ngambay, Khalkha, Upper Chehalis, Nez Perce, Yaqui, and Hawaiian, in which 'tree' and a configuration of trees, that is, 'forest,' are colexified. Furthermore, Nunggubuyu and Lengua colexify 'tree' with 'canoe,' and in Hawaiian *lāʻau* also denotes the "canoe endpiece." Maxakalí and Tsafiki colexify 'tree' and 'bone,' and in Kiliwa, *t-haq=tay* is analyzable as 'OBJ-bone=be.large' and also means 'arm, limb.'

In Rendille, Ngaanyatjarra, and Yir Yoront, terms for 'tree' are also used for 'thing' in general. Similar parallelism is found in Ngaanyatjarra and Yanomámi: in both languages, relevant terms also refer to a wooden artifact in general (a pattern also occurring in Rotuman, while the relevant Tetun term, similarly, can refer to a 'tool' or 'instrument' generally when occurring in compounds), as well as to a 'gun' or 'rifle' respectively.

Other associations include: Hausa *bishiya* is also the name of a "kind of metal helmet worn by warriors" inter alia, and Khoekhoe *haii* also means 'marihuana' in informal language. Ngambay *kake* also means 'place,' Rendille *géey* also 'dance, song,' and Muna *pughu* is also used with the meaning 'source, upholder.' Ngaanyatjarra *ngarna* also means

"entrance to honey ants' hole" and acts as a particle meaning "it's only because, it's only when" and as an adverb meaning 'temporarily;' *purnu* likewise functions as a conjunction. Sko *rí* also means 'scales of fish.' Yir Yoront *yoq* also denotes the 'Rainbow Serpent' and 'cyclone,' and *yulh*, a register-specific term, also means 'tobacco.' Badaga *mora* also denotes the "black-bark tree" specifically as well as a 'winnow, winnowing fan,' Basque *zuhaitz* can also refer to a 'tree' in Computer Science (this term is etymologically related to *zur* 'wood.' The other component is the name for the 'oak;' this etymology is similar to that proposed for an Indo-European term for 'tree'). Japanese *ki* (due to borrowing from Chinese) also means 'life-spirit' and 'intention,' Khalkha *modu(n)* also denotes the 'domino' game, and Laz colexifies 'tree' with 'milk' (accidentally, due to phonemic merger in the dialect from which the data come). Dongolese Nubian *ǧőww(ı)* denotes the 'black tree' (Acacia Arabica) specifically and Upper Chehalis *yámc* the 'Douglas Fir' specifically, *ƛ́íšaƛš* in the latter language also means 'pipestem,' and Cheyenne *hoohtsêstse* is more narrowly the name of the 'Cottonwood tree' (see Trager 1939 on this pattern). Nuuchahnulth *ƛaqaʕas* contains *ƛaq-* 'to grow.' Tuscarora *urę́ʔeh* also means 'shaft of a cart,' while Central Yup'ik *napa* as a verb means 'to stand upright' and *uqviaraq ~ uqvik ~ uqviaq ~ uqvigaq* also denotes the 'willow' specifically and is said to contain the "deep root" *uq-* 'shelter.' Cavineña colexifies 'foam,' and Guaraní *yvyra* contains *yvy* 'earth.' Jarawara *awa* is also used with the meaning 'garden' and in addition is the name of "a house made of boards in the sky where spirits are said to dwell." Kaingang *ka* also means 'mosquito,' and Miskito *dus* is also used with the meaning 'rheumatism.' Rama *kát ~ káat ~ ikát* also means 'foot,' and Wayampi *wɨla* also 'bird, dance consegrated to birds.' Bislama *tri* also means 'three' (due to collapse of Engl. *tree* and *three*), *wud* also 'carving,' and, archaically, 'penis,' and *hed* also 'head' and 'bow of ship.' Fijian *kau* also means 'to carry' inter alia, and Bwe Karen *θo* also means 'head-louse' inter alia. Hani colexifies 'tree' with stalk,' Hawaiian *lāʻau* also may refer to a 'club,' 'picture frame,' 'medicine' and conveys the meaning of 'hardness, firmness,' and presumably from there on also may refer to a "lump or knot in flesh,' 'cramp,' and 'male erection.' Samoan *lāʻau* also denotes an 'apparatus' and 'machine.' Rotuma *ʻại* colexifies 'tree, plant' with 'stiff, rigid' and other meanings, *hū* inter alia means 'lower end,' Takia *ai* also means 'pelvis,' and Tetun *hun* also 'bottom, base' and 'beginning, origin.'

66. The Valley

Represenation in Database: 77%
Motivated: 47.0%
Thereof Analyzable: 20.2% Thereof Colexifying: 27.8%
Thereof by Contiguity: 19.1% Thereof by Similarity: 6.4%
Recurrent associated meanings: gully/furrow/ditch/gorge/canal, plain/low land, river, flat/flat land, field/meadow/lawn, river bed, cave, clearing, mountain, open/opening, pampa, hole, prairie, stomach, water, bay, water course, descent

Motivated terms for 'valley' are mostly contiguity-based, although there is at least one clear recurrent pattern with metaphorical transfer realized by analyzable terms, namely

that of 'stomach' to 'valley': Khoekhoe has *!goa-!nā-b* 'depression/ditch/gully-stomach/interior-3SG.MASC,' Burarra *mu-gochila* 'CLASS.DOMESTIC-abdomen' (the term also means 'depression in the ground generally'), and Miskito *il byara* 'water abdomen.'

It is relatively rare to feature a term for 'valley' with a constituent meaning 'water' cross-linguistically. Wayampi has *ɨapɨ* /ɨ̃ɨ̃-apɨ/ 'water/river source' colexifying 'spring' and 'dew,' Hawaiian *kaha-wai* 'place-water' for both 'river' and 'valley,' Bwe Karen *lɔ̀ do* 'water flat.land,' and Lenakel *napinu*, analyzable as /napin-nu/ 'drain/gutter-water' (a semianalyzable term is found in Yuki). There are several languages in the sample in which 'valley' is associated lexically with the body of water creating valleys in the first place, namely the 'river.' 'River' or 'creek' and 'valley' are colexified in Khalkha, Nez Perce, Bora, Huambisa, and Hawaiian. Badaga *a:ru* assumes the meaning 'river' in toponyms (and otherwise also means 'to jump,' 'to get cool,' and 'six'), and a semianalyzable term of this kind is found in Kwoma. Similarly, Dongolese Nubian, Abzakh Adyghe, and Rotuman colexify 'watercourse.' Colexification of 'river' and 'valley' is also found in Arabic (Wehr 1976: 1059). Furthermore, Central Yup'ik (Nunivak Island dialect) *kuigyaneq* perhaps contains *kuig* 'river' and the postbase (see § 4.4.2.) *-neq* 'area of,' while the Lake Iliamna dialect has *kuig-na-yuq* 'river-??-thing.like.' Hani *lolxaq* contains the classifier for rivers *lol* and *xaq* 'cut open' (there is another term involving *lol*, *lolgov*, with *gov* denoting a 'thin emaciated person or animal'), One *foli sila yarole yol* appears to contain *fola* 'water, river' and *yolu* 'fall,' and Miskito *awala bak plapi tasbaya nani* contains *awala* 'river,' *bak* 'through' and *tasba* 'land.' Sentani, Abzakh Adyghe, Khalkha, Pawnee, Lengua, and Rotuman colexify 'valley' and 'river bed' (the Khalkha term is also rarely used with the meaning 'large lake' inter alia). As in Hani, terms in many sampled languages betray an association to 'open' land of some sort. Bororo has *boe-ia* 'thing-opening.' Similarly, Cahuilla *pánuwenik* contains *-pánuwen-* "'to flare out, to spread out in the full," and Kashaya *ʔama kiya·qalli* is perhaps analyzable as /ʔama kiya-qali/ 'earth extend-wide.' In Buli, Ineseño Chumash, San Mateo del Mar Huave (by the analyzable term *najmiüc iüt* /na-ajmiüc iüt/ 'AGT-fall land'), Nez Perce, Santiago Mexquititlan Otomí, Copainalá Zoque, Arabela, Embera, Kaingang, and Fijian, 'valley' is colexified with 'plain' or 'low ground, low land,' in Khalkha (by a semianalyzable term containing a verb meaning 'to blink, chop'), Chickasaw, Itzaj, Lake Miwok, and Pawnee (by a term literally translatable as "flat ground place") with 'clearing,' and in Upper Chehalis, Chickasaw, and Nez Perce with 'prairie,' in Chickasaw with 'desert,' in Bezhta with 'field' and 'village square,' in Kolyma Yukaghir with 'tundra,' and in Arabela, Cashinahua, and Tehuelche with 'pampa.' Commonly, 'valley' is also colexified with 'flat' or 'flat land,' as in Badaga, Santiago Mexquititlan Otomí, Wintu, Embera, and Kaingang, or is expressed by complex terms featuring elements with this meaning, as in the Pawnee term just mentioned, Yaqui *pa'a-la bwia* 'plain-ADJVZ land,' and Bwe Karen *lɔ̀ do* 'water flat.land' (a semianalyzable term with a constituent meaning 'flat' is furthermore found in Yuki), and in Basque, Bezhta, Lake Miwok, Nez Perce, Wappo, and Wintu, 'valley' is colexified with 'field,' 'meadow' or 'lawn' (see also § 6.3. and Buck 1949: 28 for this connection in Celtic and between cognates in Latvian and Ancient Greek). Analogously, Maxakalí *hãpxa-hit* is analyzable as 'field-remain.' Five sampled languages colexify 'valley' with 'cave.' These are Ngambay (also with 'rapidly'), Khalkha, Nivkh, Nez Perce, and Pipil.

Similarly, Buli, Kwoma and Nuuchahnulth colexify 'hole' (also "depression in ground" and 'drains' in Buli and "large hole or depression in the ground" more specifically in Kwoma). Note also that Blackfoot *sstsikómm*, which also denotes a 'coulee,' contains the verb *sstsikkii-* 'be hollow' (compare evidence from Ancient Greek reported in Buck 1949: 28 for this connection). Buli, Hausa, Baruya, Burarra, Kwoma, Kosarek Yale, Badaga, Nivkh, Khalkha, Itzaj, Nez Perce, Oneida, Wappo, Bwe Karen, Fijian, Hani, Hawaiian, Lenakel, and Rotuman colexify 'valley' with 'gully,' 'furrow,' 'ditch,' 'gorge,' or 'canal' (compare the Khoekhoe term *!goa-ǃnā-b* 'depression/ditch/gully-stomach/interior-3SG.MASC' mentioned above).

Two languages, Hani and Samoan, have terms for 'valley' that might contain elements meaning 'to separate' or 'be divided.' In Efik, the relevant term has a constituent meaning 'to descend, descent,' and these meanings are colexified in Huambisa. There are also two languages in the sample which, by spatial contiguity, colexify 'valley' with 'mountain,' namely Buin and Tsafiki. Also, there are complex terms on the basis of 'mountain': Yoruba has *ilẹ̀ lãrin òke méjì* 'earth between mountain two,' Carrier has *dzeł-î-krez* 'mountains-REL-between,' and Tetun *foho-leet* 'mountain-space' (compare the colexification of 'valley' and "interval, space between" in Piro). In addition, a semianalyzable term where one of the constituents is 'mountain' is found in Kwoma and Sedang. Takia *ilo-* also means 'inside, emotions' and 'bay,' and the latter meaning is also colexified in Hawaiian, where the relevant term is restricted to place names, and the Central Yup'ik (Yukon dialect) term *ilutak* contains *ilu* "interior, area inside; inner feelings."

Other associations include: Buli colexifies 'valley' with 'swamp,' Hausa *kwari* also means 'quiver,' and, in impolite usage, "an infant being." The Katcha term *tharɛ (na)kidhanɛ* is analyzable as 'moon middle,' Kaluli *gagodo:* may also refer to a 'dip between ridges,' and Kwoma *tabotii waw*, which can also refer to a "deep pit," contains *waw* 'deep.' Kyaka *anda*, denoting an "open valley area," may also refer to a 'hut' or 'house' inter alia. Muna *labhanga* colexifies 'vacancy, opening' as well as "place where animals frequently pass through," and *solobhangka* 'shallow valley' appears to contain *solo* 'current, flow' and *bhangka* 'boat, canoe.' Sentani *jaba* might contain *ja* 'to sink,' while Kosarek Yale *na'ob* also means 'pulse, pulse area' and, register-specifically, 'near.' The Abzakh Adyghe term *λeg°ane* contains *λeg°* 'base' and *ne* 'eye.' *T°ʔaʔ'e* is analyzable as /t°ʔe-ʔ'e/ 'twice-sink.in.' Badaga *taggu* also means "to stop, stanch, staunch, thin down the flow," and "to lower oneself, humble oneself," while Basque *bailara* colexifies "watered meadow" and also means 'borough, quarter.' Chickasaw *kochchaafokka'* is analyzable as /kochcha' aa-fokha-'/ 'outside-LOC-be.in-NMLZ,' and *yaakni' hayaka'* is literally 'earth way.off.somewhere' and denotes an 'open place' in general. Haida colexifies 'valley' with 'slough' and "insides surface or area, insides" (note that *ilutak*, a term found in the Yukon dialect of Central Yup'ik, contains *ilu* 'area inside'). The Kiliwa term *ʔ+mat=xuʔsawy* is analyzable as 'DN+earth/land=clean/clear' (with *xuʔsawy* being itself morphologically complex). Lesser Antillean Creole French *valé* also means 'to swallow, drink' inter alia (due to collapse of Fr. *vallée* with *avaler*?), while Pawnee *huukiihaar* is literally "enclosure on a surface." San Lucas Quiaviní Zapotec *ba'i* also means 'really, actually' inter alia (the reading as 'valley' is < Span. *valle*), and Guaraní *yvytyrokái* and *yvytypa'û* contain *yvy* 'earth.' Lengua *mitmegyag*

might be related to *mitmeyi* "to dig, to scratch up," and Ancash Quechua *raqra* also means 'crack, split.' Fijian *qākilo* seems to contain *kilo* "a low place, ravine, hollow in centre of whirlpool or water being stirred in a cup" (also meaning 'low, hollow, depressed' adjectivally), and *qiloqilo* is reduplicated from *qilo* 'hollow in tree where water stagnates.' Hawaiian *awaawa ~ awāwa* 'valley, gulch, ravine' appears to be reduplicated from *awa*, meaning 'harbor, cove' and 'channel, passage through reef.' *Kuawa*, a poetic term for 'valley' in the same language also means 'guava' due to English influence, and Malagasy *lohasàha* is analyzable as 'head-rice.field.'

67. The Volcano

Representation: 20.95%
Motivated: 52.67%
Thereof Analyzable: 97.97% Thereof Colexifying: 2.03%
Thereof by Contiguity: 70.92% Thereof by Similarity: 26.02%
Recurrent associated meanings: burn/fire, mountain/hill, earth/land, melt, explode

Frequently, complex terms for the 'volcano' are of the lexical type and minimally consist of elements meaning 'mountain, hill' and 'fire, fiery' or 'burn,' as in Buin *menu oguai* 'mountain fire,' and at times also further constituents, as in Nivkh *t'uɣr-kir t'a bal* 'fire-INSTRUMENTAL breathe mountain.' Such terms are also found in Yoruba, Kyaka (which also has an alternative term where 'lightning' instead of 'fire' figures as one of the other elements alongside 'mountain'), Basque, Khalkha, Welsh (where *llosg* means, alongside 'burning,' also "arson, inflammation, scald"), Xicotepec de Juárez Totonac, Guaraní, Miskito, and Samoan.[7] Alternatively, in some languages also 'land' or 'earth' rather than 'mountain' is featured in complex terms as in Efik, where *ayua obüt ikañ* contains *obüt* 'land' and *ikañ* 'fire,' Guaraní *yvy-rata* 'earth-fire' (Guaraní also has *tata-po* 'fire-hand,' *tata-sê* 'fire-come.out, and *tata gue'êha*, containing *tata* 'fire' and *gue* 'to extinguish'), Bora, where *tsátsihdyu ííñújɨ pañétú cúújuwa íjchívyéne* contains *ííñújɨ* 'earth' and *cúújuwa* 'fire,' Hawaiian *ahi ʻai honua* 'fire consume land,' and Tetun *rai-suut* 'earth-gap.' Another complex term involving 'fire' is found in Baruya, which has *wa'nɨ-dɨka* 'permanent-fire' (this term indeed also denotes a "fire that has been burning for a long time").

Moreover, in two sampled languages, Chickasaw and Chayahuita, terms for 'volcano' contain an element meaning 'to melt': *tobaksi' bila-'* 'coal melt-NMLZ' (this term also denotes 'lava'), and *na'pi soquirin-so' no'pa quëran pashí tënin-so'* 'stone melt-3SG.SUB earth from ?? say-3SG.SUB' respectively. Also in two sampled languages, relevant terms contain a constituent meaning 'to explode': Wintu *pʰuyuq pʰuqa·* 'mountain explode/erupt' and Arabela *jiya taaniu* 'earth explode.'

Other lexico-semantic associations are few: Itzaj has *aj-noj witz* 'MASC-big mountain' (this term is glossed also as 'mountain' itself), and Hawaiian *pele* also means 'eruption,' 'lava flow,' and 'soft, swollen, fat,' and is also the name of a volcano goddess.

[7] Japanese has this pattern, too, but the consulted source does not mention the relevant term and hence this case is outside the present sample.

68. The Waterfall

Representation: 67%
Motivated: 52.0%
Thereof Analyzable: 38.4% Thereof Colexifying: 13.6%
Thereof by Contiguity: 31.5% Thereof by Similarity: 6.6%
Recurrent associated meanings: water, current/rapids/cataract, fall, cliff/precipice, river, jump, flow, down

English *waterfall* is a representative of the cross-linguistically most common structure of analyzable terms for the 'waterfall,' featuring consituents meaning 'water' and 'fall.' In the sample, they are found in Khoekhoe, Noni, Basque, Welsh (in addition, there is another term which directly colexifies 'waterfall' with 'flow, fall' in this language, and Wintu *tEk* likewise colexifies 'waterfall' with 'flow,' next to 'extrude, be extruded' and 'press'), Chickasaw, San Mateo del Mar Huave, Itzaj (colexifying 'jet of water'), Kashaya, Santiago Mexquititlan Otomí, Yaqui, and Sáliba. Moreover, for Haida *quuga*, compare *quugaa* 'have fall.' The source notes: "*quuga* may in fact have originated as a nominalization of the stative [verb *quugaa*] rather than the latter being derived from a noun." There are also many languages with terms with a similar structure, but where the semantics of the constituents is slightly divergent. Yoruba has *ọ̀ṣọ̀rọ̀ omi* 'pour.down water' (note in this context that Buin *parukuna* is derived from a verb meaning 'to spout, to pour'), Sora *tuŋsar'da:n*, containing *t^{ʔ}u:ŋ-* "to collapse, drop down, pierce" and *d^{ʔ}a:-* 'water,' Oneida *tetwaʔsʌ́thaʔ*, analyzable as /te-w-aʔsʌht-haʔ/ 'DUALIC-NEUT.AGENT-drop:CISLOCATIVE-HAB,' Pawnee *kictakaahak*, literally "water to pass down," which is a verb meaning 'be a waterfall' as well as 'water to drip' and 'be a rapid' and contains elements meaning 'be liquid' and 'down,' Xicotepec de Juárez Totonac *yujyā xcān*, presumably /yujā xcān/ 'move.down water,' Yana *ba-riʔmau-xa* 'spill/flow-place-water' (this term also denotes a particular site; note also that in Carrier there is a semianalyzable term for 'waterfall' containing a verb 'to flow' and compare also Cheyenne *anôhehéeóó'e*, which is analyzable as /anôhe-hée'o'tsé/ 'down-spill' and also means 'to pour down' when interpreted verbally), Takia has *you i-skalik da* 'water 3SG-pour.away IPFV,' Tetun *beetudak* containing *bee* 'water' and *tuda* 'hurl, fling,' and Tuscarora has *yuhtawę́ʔę*, containing *-htą́w-* 'stream of water, current of water' and *-ę-* 'fall.' Chukchi has *emlaratɣərɣən*, containing *eret-* 'fall' (subject to vowel harmony), and Rama *kála taik* 'fallen nose.'

As in Tuscarora, there are also terms in which 'river' rather than 'water' alongside 'fall' figures. These include Kapingamarangi *monowai doo* 'river fall' and White Hmong *dej poob tsag* 'river fall cliff.' Similarly, Yay has *ram^{6} tok^{3} taat^{2}* 'water fall cliff' as well as *taat^{2} ram^{6}* 'water cliff,' where *taat^{2}* is "cliff over which water falls," "cliff which is the site of a waterfall" (a semianalyzable term including 'river' is found in Embera). In line with the Southeast Asian association with 'cliff' by overt terms, Bwe Karen colexifies 'waterfall' and 'steep cliff, precipice,' inter alia. Similar to the association with 'cliff' in the Southeast Asian languages is also Copainalá Zoque *nʌ'ʌtʌŋgomʌ'nguy* /nʌ'-tʌŋgomʌ'nguy/ 'water-precipice.'

In Yoruba, Hani, and Hawaiian, there are terms for 'waterfall' containing verbs referring to some sort of jumping motion: Yoruba has *ì-takiti omi* 'NMLZ-to.somersault water,' Hani *(wulquvq) coq-ceiv* '(water) jump-break' alongside *wulquvq coq-ciivq* 'water jump-pinch/choke,' and Hawaiian *wai-lele* 'water-jump/leap.' Other complex terms for 'waterfall' with 'water' being one of the constituents include Meyah *mei ofoská* containing *mei* 'water, river, sperm' and presumably *ofos* 'skin, mountain top,' Sora *taŋ'surda:n* /taŋ'surd^{ʔ}a:-n/ 'rise-water-N.SFX,' Guaraní *y-tororõ* 'water-gush' and *y-tu*, probably analyzable as 'water-father,' Piro *gijrotga* /gijrota-giga/ 'forehead-water,' Wichí *inot n'oyij* 'water way,' and Great Andamanese *înal'ârchâr*, containing *îna* 'fresh water' and *châr* 'spring of water.' Semianalyzable terms for 'waterfall' on the basis of 'water' are found in Berik, Kyaka, One, Yir Yoront, Nivkh, Bislama, and Lenakel. In Kiowa, 'waterfall' is *zout-syᵸn-goup* 'current-small-hit' and in Ket it is *qaʁɯn* /qa^{ʔ}-quūn/ 'big-current.' 'Current,' 'rapids,' or 'cataract' are colexified in Khoekhoe, Kwoma, Khalkha, Nivkh, Upper Chehalis, Itzaj, Pawnee, Quileute, Aguaruna, Bora, Chayahuita, Cubeo, Hupda, Miskito, Ancash Quechua, Yanomámi, Tetun, and Hawaiian.

Other associations include: Ngambay *kóró* also means 'storm' and denotes a kind of snake, and Swahili *maporomoko* consists of the verb *poromoko* "to slide down in a mass" and the noun class prefix *ma-*. Kyaka *jingi* also means 'flower, bloom.' Muna *kasapa* is derived from the verb *sapa* 'to splash,' and Rotokas colexifies 'waterfall' and 'steep slope' as well as 'stone face.' Kosarek Yale *modun* 'place of a waterfall' also means "upper edge of a landslide." The Bezhta term *ǯišiyo* contains a verb meaning 'to entangle' and the past perfective participle suffix. Blackfoot *ohpsskonaka'si* also means "to spurt/flow rapidly" as well as 'geyser.' Nez Perce *tíkem* also means 'dam' and 'fish bladder.' Nuuchahnulth *tuxʷiɬ* also can refer to the 'Sproat Lake falls' and "the falls on the Sarita river," and Oneida *tetwaʔsáthaʔ* also to the 'Niagara falls' specifically. Central Yup'ik *qurrlugtaq* contains *qurre-* 'to urinate; to spawn (of fish).' Bora *waapéwa* colexifies 'spout, stream' and contains *waapé-* 'to drip' and the classifier *-wa* for table-like entities, and *nééwabya* contains *nééwa* 'stone' and the classifier for three-dimensional objects *-bya*, while Miskito colexifies 'waterfall' with 'rock, big pebble in the middle of river.' Bwe Karen *bla* also means 'to wash the face,' and Samoan *āfu* also means '(to) sweat,' '(be) heated,' and 'to wither.'

69. The Wave

Representation: 70%
Motivated: 19.8%
Thereof Analyzable: 17.5% Thereof Colexifying: 2.3%
Thereof by Contiguity: 2.9% Thereof by Similarity: 14.9%
Recurrent associated meanings: water, river, swell/swelling, sea, wave (techn.), skin/surface, stir, gust, surge

Motivated terms for the 'wave' ('billow,' 'ripple,' 'breaker,' 'surf') are relatively rare cross-linguistically. Where they occur, they are likely to be similarity-based. In many languages, terms feature verbs referring to the motion of the waves, a situation also found in Indo-European languages (Buck 1949: 40). Wayampi *palanayaapiyapi* is analyzable as /palana-y-

apiyapi/ 'sea-NMLZ/REFLX-to.surge,' and Abzakh Adyghe *we-λe* as 'surge.up-foot.' Baruya *mavajɨnaaka* colexifies 'wave' with 'swirl,' Muna *kaendo* is derived from *endo* 'to shake' and *lolebhata* "calm waves without foam" might be related to *lole* 'to roll around.' Kosarek Yale *mak kola'ebna* contains *mak* 'water' and a verb meaning "to make stirring movements" (not necessarily of water, though), and Hawaiian *nalu ~ nanu* colexifies 'wave' and 'to stir' inter alia. Kyaka *ipwua ~ ipya kaso kole minyingi* contains *ipwua ~ ipya* 'water' and *kaso kole* 'up and down,' (with *kaso kole minyingi* glossed as 'wave movement'), Biloxi has *ani´ xoxo-ni´* /ani´ xoxo´-ni´/ 'water swing-CAUS,' Cheyenne *-nêšêške'sevó* contains *nêšê'* 'to wash' and *'sevó* 'to flow,' a constituent of San Mateo del Mar Huave *ateapteapeay* is *ateaptea-* 'to move by wind,' Santiago Mexquititlan Otomí *munts'i* is analyzable as /m-punts'i/ 'NMLZ-to.relapse,' Yaqui *bawero'aktim* as /bawe-roa-k-te-im/ 'sea-turn-PFV-INTR-PL,' Wintu *le(·)w* is connected to a root meaning 'vibrate, oscillate,' Bororo *pobo amagadu* is analyzable as 'water agitation,' and Guaraní y-*pu'ã* is analyzable as 'water-rise.up.' Abipón *lilikatka* contains *ili(k)-* 'movement of water,' and finally, Dongolese Nubian *árrɛ* goes back to **ánrɛ*, containing *án-* 'to go.'

However, there is also a wealth of terms for 'wave' in the sample that make reference to the shape rather than the movement of waves, and these are frequently metaphorical in nature. Specifically, there are a lot of body-part metaphors (as already seen in the Abzakh Adyghe term mentioned above): Yir Yoront has *yuwl-man* 'sea-neck,' Carrier *tha-tši* 'water-head,' and Bororo *iagajaga* and *iagiri* seem to contain *ia* 'mouth.' Other metaphorical denominations include: Hausa *rak'umi-n ruwa* contains *rak'umi* 'camel' and *ruwa* 'water' (*rak'umi* can also refer to the "blossom of locust-bean tree before it is fully out"), Kanuri *kàzáà njî-bè* is analyzable as 'spear water-of,' Bezhta colexifies 'wave' with 'horse,' and Pawnee *kicpiirakus* is analyzable as /kic-wiira-kus/ 'be.liquid-upright-be.sitting.' The Piro verb *hawokhata* 'to be turbulent, have waves' contains *hawoka* 'to blow,' and *špurha* contains *špu* 'edge.' Yay *ram*^{6} *foŋ*^{4} is presumably analyzable as 'water roof,' and *me*^{5} *foŋ*^{4} as 'mother roof.' Guaraní also has *y apeno* containing *y* 'water' and *ape* 'skin, surface,' and the latter two meanings are associated with 'wave' by direct colexification also in Welsh, while there is a semianalyzable term in Central Yup'ik.

As seen from the data discussed so far, frequently complex terms for 'wave' unsurprisingly contain a constituent meaning 'water.' Further terms of this structure are Kwoma *uku veereveer* 'water breeze/light.wind,' Bororo *pobu-to* 'water-inside,' and Guaraní *y joapy* 'water connected.' Rama *albríni* in *sí albrínima* 'wave' means "waving, grinding" (*sí* is 'water'). Semianalyzable terms with 'water' are found in Kwoma, Kosarek Yale, Cavineña, and Guaraní, while Ngambay directly colexifies 'wave' with 'water' as well as 'river' and 'well;' Sko and Lenakel have semianalyzable terms where one of the constituents can be identified to mean 'sea.'

Furthermore, Hawaiian *'ale* rarely means 'gust' inter alia, and White Hmong has the analogous complex term *nthwv dej* 'gust water.' In Efik, the 'wave' is *mbufut akpa* 'swelling river,' for which compare Berik *fo buk* 'wave' with *bukna* 'to swell' (*fo* is 'water, river, lake'); Basque colexifies the meanings, and the association is also present in Tasmanian according to Plomley (1976: 380) and in Indo-European languages (evidenced for instance

in Latin and Ancient Greek, Buck 1949: 40). In Basque and Mandarin, the relevant terms can inter alia also refer to a 'wave' in the technical sense, e.g. a 'radio wave.'

Other associations include: Rendille colexifies 'wave' with 'storm' on a lake or the sea, Swahili with "bulrush millet, eleusine," and Kwoma *uku woya* also denotes a "decorative zigzag band on clay pots and other artefacts." Kyaka *pyakao pulyuo lupyuo pingi* contains *pulyuo* 'upwards' and *lupyuo* 'down, descending.' Toaripi *aroaro* also denotes 'charcoal,' probably by extension from there "the Frigate Bird; black with white breast," as well as a 'squid' and the *Mimosa pudica*. Sahu *moku* also means 'dizzy.' Basque *itsaski* contains *itsas* 'sea' and also denotes 'seafood,' while *olatu* as a verb means 'to tame, domesticate' and 'to beat, pound.' Khalkha *kyrkire*, as a verb, also means 'to growl, grunt, snarl,' *calgija* contains the verb *calgi-* 'to splash' and *dolgija(n)* is related to *dolgi-*, meaning 'to wave, undulate,' but also "to be restless, tempestuous, irascible." Welsh *ton* also means 'lay-land' and 'broken.' Blackfoot *ohpai'kímsskaa* may contain *ohpai'piiyi* 'to jump,' Kashaya *daluw* also means "rub or spread something by hand," Chickasaw *bo'kalhchi* 'to come in waves' also means "to be beaten (of eggs), to be splashed," Lesser Antillean Creole French *lanm* also means 'blade,' Itzaj *kukul* is a verbal noun derived from *kul* 'to roll,' and Nuuchahnulth *kʷax* also means 'spray.' Central Yup'ik *yuulraaq* also denotes a 'thin flexible sheet of ice on ocean.' Bora *tyocáhco* also denotes the sound of water splashing against something, Cayapa *sela* is also the term for the 'sky.' Hani *eeldaol*, *eeldaol bi*, and *eeldaol hhev* contain *daol*, meaning 'to exist, lead the way' and 'one line' inter alia; *eel* is glossed as 'to laugh, to smile,' but occurs in many complex terms for water-related meanings. Quaintly, *hhev* in the last variant means 'to wave one's hand, to gesture.' Fijian *ua* 'wave, tide' also denotes 'veins, tendons' and 'muscles.' Hawaiian *nalu* ~ *nanu* also inter alia means 'to ponder, to reflect' and *'ale* also 'for tears to well in the eye.' Malagasy *àlona* also means 'jealousy,' Rotuman *vạlu* also 'eight,' Samoan *galu*, as a verb, also 'for the sea to be rough,' and White Hmong *twv* also 'to dare' inter alia.

70. The Wax

Representation: 55%
Motivated: 45.7%
Thereof Analyzable: 24.5% Thereof Colexifying: 21.2%
Thereof by Contiguity: 15.2% Thereof by Similarity: 25.5%
Recurrent associated meanings: candle, honey, bee, resin, fat/grease, faeces, tar, ear-wax, house

Many denominations for the 'wax' are metaphorical in nature cross-linguistically, with or without 'bee' acting as contiguity anchor. Most frequent, however, is a pattern of colexification by provenience contiguity, namely that with 'candle,' found in Khoekhoe, Dongolese Nubian, Muna, Greek, Khalkha (the relevant term is a borrowing from Chinese in this sense; it has also other indigenous meanings), Welsh, Highland Chontal, Itzaj, Xicotepec de Juárez Totonac, and Tetun (in the latter languages, from 'candle' extended also to 'lamp'). Furthermore, Chayahuita has *pa'nan*, where *pana* is 'side of the candle' and *-nan* an instrumental suffix. Taken together, however, metaphor-based associations with substances

which are, like 'wax,' semi-solid, formable and/or have a similar texture to the target concept are more frequent. For instance, the Khoekhoe term exhibiting this pattern of colexification, *ǂnuru-llnui-b*, is analyzable as 'knead-fat/oil-3SG.MASC.' In fact, 'fat' and 'wax' are colexified in Samoan (alongside "medicinal cream, wax"), and Imbabura Quechua *mishki wira* is analyzable as 'bee grease,' while Abzakh Adyghe *śef* may contain *śe* 'fat' and *fə* 'white.' Further, Efik has *a'dan u'töñ* 'oil grease ear' (compare *utön unam* 'glue' "[s]o called from being supposed to be made of the ears of animals"), and there is a semianalyzable term where a constituent may mean 'greasy' in Hawaiian. In Tuscarora, Hawaiian, and Tetun, 'wax' and 'resin' are colexified (Tuscarora also colexifies 'cement,' 'glue,' 'gum,' 'jelly,' and 'syrup' by the same term, and Tetun also 'candlenut' and 'to prune, clip'), while in Wintu the association is realized by the analyzable term *hub-in ceki* 'bee-LOC pitch.resin.' Similarly, Haida colexifies 'wax' with 'pitch, tar' (and also 'chewing gum'), Tuscarora also colexifies 'wax,' 'tar,' and other meanings, and Kildin Saami has *veazvušk-tārr'v* 'wasp-tar.' In Buin, 'wax' and 'faeces' (and also 'fine powder' generally) are colexified, and this association is mirrored by complex terms in Buli (*si-beung* 'fill.up-faeces/droppings'), Carib (*wano weti* 'bee excrement'), and Tsafiki (*chiná pe* 'bee excrement; compare Cayapa *chiñape,* where the general term for 'bee' is, however, *tanda*, and *chi'ñilla* is a term for a type of bee similar in appearance to a wasp). There are also many languages in which 'wax' is associated with 'honey.' Herein, Buli is the only language with colexification (though Nunggubuyu has a broad term that may include reference to 'wax' as well, although not explicitly glossed so), while in other languages the association is realized formally by complex terms: Kanuri has *kəmàgən-mí* 'honey-son.of,' Kiliwa *mi?yawy=smaq* 'honey=leaf,' Guaraní *eiraity* /eíra-ty/ 'honey-urine/juice,' Hupda *nɛ̌ŋ cak* 'honey mass,' Toba *pic l-apa* 'honey 3SG.POSS-wasp.nest/honeycomb,' and Great Andamanese *âja-pîj* 'honey-hair/feather.' Miskito and Rama colexify 'wax' with 'ear wax' (Miskito also with 'scab' and 'scabies'). Semianalyzable terms involving 'honey' are featured in Hausa (where 'honey' is colexified with 'bee'), Guaraní, and Hani. Moreover, two languages, Yaqui and Fijian, have terms with a metaphorical transfer from 'house' to 'wax': *muumu jo'ara* 'bee house' and *drega ni vale ni vī*, containing *drega* 'gum' and *vale* 'house' respectively. Other complex terms with 'bee' are Chickasaw *fohi' bila-'* 'bee/honey melt-NMLZ' and Tetun *bani-isin* 'bee-flesh,' and semianalyzable terms with 'bee' are found in Hausa, Katcha and Huambisa (for the unknown element *nujiri* in this language, compare *nujin* 'egg'?).

Hausa *kaki* also means 'mucus' inter alia, Muna *lili* also to "go or travel around," and Basque *zira* also 'shoe polish.' Noni *kelay* is formed by adding the noun class prefix *ke-* to *lay*, which means 'to seal, glue.' Gurindji *tarla* also denotes "wax" from the Spinifex plant, Badaga *mekku* also means 'cud,' Lesser Antillean Creole Frensh *si* also "sour, having an acid taste" inter alia, and Pipil (Cuisnahuat dialect) *cha:pah* is also the name of a game played with wax. Aguaruna *dují ~ nují* also means 'nose, beak' and 'prow of canoe,' Bororo *bori* also denotes the bony scales of an alligator, and Sedang *pet* also means 'to plant' inter alia.

71. The Whirlpool

Representation: 55%
Thereof Motivated: 50.1%
Thereof Analyzable: 43.1% Thereof Colexifying: 7.4%
Thereof by Contiguity: 29.2% Thereof by Similarity: 19.2%
Recurrent associated meanings: water, go around/spin/turn/twist, whirlwind, eye, vessel, current, sea

Most frequently, terms for 'whirlpool' ('eddy,' 'vortex,' 'maelstrom') are morphologically complex in the languages of the sample, containing terms for 'water' or specific bodies of water and verbs for motions such as 'to go around,' 'to spin,' 'to turn,' and 'to twist,' as in Koyraboro Senni *hari-windi* 'water-go.around/go.in.circles.' Other languages with terms with such a structure and varying water-related contiguity anchors are Bezhta, Ket (which also has the alternative term *telɣil* /tel-kil/ 'mammoth-spin,' the analysis of which is unsure however according to lexicographers), Welsh, Ineseño Chumash, Haida, Kashaya, Kiliwa (*xa?-t s+wa-y-p=s+wa-y-p* 'water-SUBJ INST.LONG+spiral.motion-ATT-PASS=INST.LONG+spiral.motion-ATT-PASS), Oneida, Yaqui, Guaraní, Toba (where the element conveying the motion is 'to return' more specifically), Hawaiian, Malagasy (where the same remark as for Toba applies), White Hmong, Samoan, and Vietnamese (where an additional element meaning 'current' is present, compare Biloxi *aní´xyu´hi kĭdu´nahí* 'current turn.around' and Haida *juu rwiihlrahl*, containing *juu* 'current,' *rwiihl* 'to move spirally' and *rahl* 'to stir;' this term also denotes a 'crow's nest at the back of head'). Similar terms of the derived type are encountered in Piro, Wayampi, and Hawaiian (by the term *mimilo*, which is formally a reduplication of *milo* 'twist, whirl' and also means 'curly, kinky' as well as "to roll, as to induce an abortion"). The relevant meanings are colexified directly in Kyaka (although the situation in this language is not entirely straightforward on the basis of the consulted source), Sora, and diachronically, such an association is also identifiable for a Kolyma Yukaghir term for 'whirlpool.'

There are, however, also metaphor-based conceptualizations of 'whirlpool.' In four sampled languages, there are complex terms in which 'eye' is transferred to 'whirlpool,' with 'water' or bodies of water acting as a contiguity anchor: Kanuri has *shim njî-bè* 'eye water-of,' One has *fola namna toma* 'water eye stone,' San Mateo del Mar Huave *oniiüg ndec* 'eye sea' for a whirlpool in the sea specifically, and Takia *you mala-n* 'water eye-3SG.' In three of the languages in the sample, the source concept is a vessel: Nez Perce *capahiƙayiƙáyica* 'funnel-shaped whirlpool' contains *cepé·* 'to shape with hand' and a reduplicated version of *hi·ƙay* 'pot, cup, to make a cuplike shape,' and in Bororo, the 'whirlpool' is *aria-reu* 'pot-like.' In addition, Hupda *wɔ̃wɔ̃´y'* colexifies 'whirlpool' with a "basket type having hourglass shape" and "gourd support (hourglass shape)." Basque (by the term *zurrunbilo*, perhaps related to *zurrun* 'pole' and *bilo* 'down, fine hair'? This term also colexifies 'bedlam,' 'confusion,' 'crowd,' and "heavy downpour"), Cubeo, and Fijian colexify 'whirlpool' with 'whirlwind.' Since the Fijian term is morphologically complex and contains *waitā* "water deep enough for a canoe at low tide," it appears that 'whirlpool' is the diachronically original referent of the term. The contrary situation holds in Santiago

Mexquititlan Otomí, where the 'whirlpool' is called *xedi dehe* 'whirlwind water,' and the same is true of Abzakh Adyghe and Imbabura Quechua.

Other complex terms with 'water' include Toaripi *ma elorielori oti* containing *ma* 'water' and *oti* 'place,' Kildin Saami *čāʒ'-čullm* 'water-knot,' Wintu *meˑm kumuˑri*, containing *meˑm* 'water, river' and *kum*, referring to the 'sound of water rushing or roaring' and also meaning 'water turbulence,' Bororo *buto-bo*, seemingly 'rain-water,' and Cavineña *ena cahuaitiya* 'water enrage.' Semianalyzable terms where one of the constituents can be identified to mean 'water' are found in Highland Chontal, Aymara, and Hani.

Still other associations include the following: Efik *ësiḱ* is derived from the verb *siḱ* 'to squirt;' the term also denotes a "round thing drawing in towards the centre" and a "round net or fishing pot," a pattern of colexification probably due to the fact that *siḱ* also means 'to tie, draw together' inter alia. Khoekhoe colexifes 'whirlpool' with 'pot-hole' in river and 'gravel pit' inter alia, while Muna *tehi tingkulu* is analyzable as 'sea slope.down' and also denotes the ocean itself when no land is in sight. Greek *roufx̌chtra* contains *rouf-* 'to suck in, absorb,' Khalkha *oil* also can refer to a 'tuft of hair,' and Kolyma Yukaghir *unuŋ-jurugu* 'eddy in a river' is analyzable as 'river-hole.' Tuscarora *nekahtawakwá?nahč* contains *-htg̲w-* 'stream of water, current of water' and *-kwa?n-* 'arc, curve.' Cayapa colexifies 'whirlpool' with 'grindstone,' Cubeo *jõmeicobe* is derived from *jõmeñu* 'to agitate' by means of the classifying suffix *-cobe* for hole-like objects, and Guaraní *jepyvu* contains *yvu* 'spring.' Piro *ximlere* apparently contains *ximle* 'to boil,' for which compare Central Yup'ik *qalla-neq*, colexifying 'spring' with 'eddy' and analyzable as 'be.boiling-thing.that.results.from.' Ancash Quechua *muyu* also denotes a 'rodeo' or 'walking around' (Spanish gloss: 'rodeo'). Samoan *auma* also denotes a 'breaker,' and Tetun *da-dula-k*, analyzable as 'DERIV-winnow/sift-DERIV,' also means 'to sieve for grain.'

72. The Airplane

Representation: 65%
Motivated: 55.5%
Thereof Analyzable: 50.2% Thereof Colexifying: 5.2%
Thereof by Contiguity: 4.7% Thereof by Similarity: 28.8%
Recurrent associated meanings: fly, boat/canoe, bird, sky, machine, car/vehicle, air, high, metal, air, helicopter, wing, house, jump, thing, travel

Most frequently, complex terms for 'airplane' are derived from verbs meaning 'to fly' or 'to fly around,' with a variety of structural subtypes. Derivation proper is attested as the word-formation device in Gurindji, Ket (where the relevant term also can refer to a helicopter; the same is true of a simplex term in Sko), Carrier, Kashaya, Nuuchahnulth, Oneida, Pawnee, dialectal Central Yup'ik, Aguaruna, Guaraní, and Wichí, which has *wiy'o-taj* 'fly-AUG.' In two languages, additional elements are present, namely Nez Perce (*weˑkeʔykeʔí* /weˑ-keʔéy-eʔí/ 'fly/run-move-AGT') and Yanomámi (*ɨsɨhami-yë-rewë* 'up.high-fly-NMLZ'), and in another two sampled languages, there are complex terms that are somewhat akin to derivation with an element meaning 'thing,' namely Cheyenne (*ame'hahtôtse*, analyzable as /ame'há-hestôtse/, 'fly.along-thing') and Hupda (*wayd'ʊ́ʔ-teg*

'fly-thing'). In two sampled languages, there is a metaphorical transfer from 'house' to 'airplane' with 'fly' acting as a contiguity anchor. These are Rotokas (*papa-pa kepa* 'fly.through.air-DERIV house/cabin/building') and Kiliwa (*ʔ-waʔ=kw-iʔ+hiw* 'DN-house=WH-DIST+fly'). In Kashaya and Yanomámi, there are additional constituents in the relevant terms meaning 'high' (*qalicatʰmuʔ* /qali-hca-tʰ-mu-ʔ/ 'high-fly-PLURAL.ACT-around-ABS' and *ɨsɨhami-yë-rewë* 'up.high-fly-NMLZ'); 'high' is also featured in Bora *cáámé-e-mɨ* 'be/become.high-belong.to-SCM.transport' and Piro *tenyapatš̲ro,* containing *teno* 'high, tall, deep' and *ya* 'to go.' In Santiago Mexquititlan Otomí, 'metal' is present as the meaning of the second constituent of the respective 'airplane'-terms alongside 'to fly': *nsani bojä* 'fly iron/car,' while Bororo has *meriri kodureu*, containing *meriri* 'metal' alongside a verb meaning 'to fly' and White Hmong has *dav hlao* 'hawk iron.' Semianalyzable terms with 'fly' are featured in Basque, Khalkha (in varieties spoken in inner Mongolia), Lakhota, and one on the basis of a verb meaning 'to fly alone' in Toba. It is also not uncommon to transfer 'car,' or more commonly 'vehicle' in general, to 'airplane,' similar to Santiago Mexquititlan Otomí: Kanuri has *mààrá sámí-bè* 'vehicle sky-of,' Khoekhoe *ǂoa-kuni-s* 'air-wagon-3SG.FEM,' Yoruba *ọkọ̀ òfurufú* 'vehicle sky,' Sora *'ruaːŋ-səgada-* 'sky-cart-,' Cashinahua *nai bapu* 'sky motor/motored vehicle,' Malagasy *fiaramanìdina* /fiàra-manìdina/ 'vehicle fly,' and Mandarin *fei1-ji1* 'fly-machine,' for which compare Japanese *hi-kō-ki* 'fly-go-machine' and Vietnamese *phi cơ* 'fly machine' and *máy bay* 'machine fly.' In the Americas, there are three languages, Kashaya, Aguaruna, and Huambisa, which colexify 'machine' or 'machine with motor' in general with 'airplane.' In all instances, the relevant terms are borrowed from Spanish, from *vapor* in the case of Cashinahua (and note that Chayahuita *huaporo*, also a borrowing from Spanish, colexifies 'steam' with 'airplane') and from *machina* in the rest of the languages. Kashaya has the optional complex term *ma·kina ca·dmuli*, which is tentatively analyzable as /ma·kina hca-vᴼd-mul=li/ 'machinery fly-along-around=INSTR.' More frequent than associations with 'car,' 'vehicle,' or 'machine' is the transfer from 'boat' to 'airplane,' realized almost exclusively (with the exception of Khalkha which has a colexifying term that is indeed parallel to English *vessel* in being also capable of referring to a receptable or trough) by morphologically complex expressions, with either 'sky' as contiguity anchor, as in Hausa *jirgi-n sama* 'boat/train-GEN sky,' or more frequently with 'fly,' as in Rotuman *'ahại fere* 'ship fly' and also in Muna, Sko, Meyah, Nivkh, Chickasaw (where 'boat' is colexified with 'trough'), Cayapa, Fijian, and Hawaiian ('canoe' rather than 'boat' specifically in the case of Cayapa and Fijian), Rotuman, Samoan, and Yay, with 'jump' in Abzakh Adyghe *qᵒʔəḥλate* /qᵒʔəḥe-λete/ 'boat-jump,' and with 'air' in Bwe Karen (*kəbɔ́-gəli* 'ship-air') and Hawaiian (*moku-ea* 'ship-air').

All of the second elements in the above terms recur in other configurations in other languages. Like Hausa, the meaning 'sky' figures in Ngaanyatjarra, where *yilkaringkatja* contains *yilkari* 'sky' and *katja* 'son' (this term also means 'airline, air services'), and Lengua *thlinga netin* 'movement sky.' In Miskito, 'airplane' is *pasara pali tauki ba* 'air jump travel DEM,' terms of the derived type involving the meaning 'fly' were already discussed above. In Guaraní, 'wing' and 'airplane' are colexified and there are further semianalyzable terms containing *pepo* 'wing.' Note also in this context Yir Yoront *minh-puth lon* 'animal-wing with,' but this is due to another metaphorical transfer pattern: *minh-*

puth lon in fact is the Yir Yoront term for 'bird' (compare section 6), and with this pattern of colexification, Yir Yoront is accompanied by Swahili, Toba, and Mali, where 'large bird' specifically is colexified wih 'airplane.' Khoekhoe has *anis kunis* 'bird wagon,' Ngambay *yèel bò* 'bird big,' and One *tolla moa* is presumably analyzable as /tolla mo'a/ 'bird mother' (with 'mother' presumably conveying augmentative meaning, Matisoff 1992). Bakueri *mẹ́lí mẹ́ ngoú ~ mẹ̱li mẹ́ ngoú* contains *mẹ̱li,* which denotes a species of bird, and *ngoú* 'fever, cold.' As for 'bird,' there is also a semianalyzable term in Guaraní.

There are also terms for the 'airplane' in which 'air' in fact is a constituent: alongside the terms already discussed, this is also the case in Welsh, where *awyren* contains *awyr* 'air, sky' alongside a singulative suffix. Semianalyzable terms including elements with this meaning also occur in Greek (although *aeroplánon* is in fact borrowed from French, which is in turn a learned neologism based on the Ancient Greek root for 'air'), Lengua, and Manange (where 'air' and 'wind' are colexified).

Other more unusual patterns include: Cubeo *vɯicũ* contains the root *vɯi* also occurring in terms for 'propeller,' 'lungs,' and 'dry tuber' and a classifier for things that are round on the one side and plane on the other. Guaraní *kurusu veve* contains *kurusu* 'cross,' and Bislama *plen* (< Engl. *plane*) colexifies 'wood plane.'

73. The Ball

Representation: 74%
Motiveted: 31.7%
Thereof Analyzable: 12.5% Therof Colexifying: 19.3%
Thereof by Contiguity: 15.5% Thereof by Similarity: 7.6%
Recurrent associated meanings: round, bullet, football, baseball, citrus, lump, play ball game, testicles, nut of dum-palm, bowl, kick, rubber

The most common of the few lexico-semantic associations with 'ball' (ignoring additional glosses such as 'sphere, spheroid') is that with 'round, round thing' (see Buck 1949: 907 for evidence from various Indo-European languages). It occurs either by morphologically complex terms, as in Meyah *móf ofóg* 'wind round,' One *malwa tala ~ maula rala* /malwa tala/ 'citrus round.thing,' Nez Perce *ƛ̓apapƛ̓ápap*, reduplicated from *ƛ̓ápap* 'to be round,' and Lengua *aksak yakye* (*aksak* is 'thing' and *yakye* likely related to the verb *yakyeyi* 'to be round') and directly by colexification in Buin, Rotokas (where the relevant term *kororoisia* appears to contain *koro* 'fruit'), Khalkha (one of the relevant terms, *bømbyrceg*, seems to be related to *bømbyr* 'drum'), Santiago Mexquititlan Otomí (where the relevant term also means 'wheel'), Quileute, Wintu (where the relevant term also means 'naked'), Embera, Hupda (where the relevant term also means 'fruit' and 'nut'), and Piro (where, similarly, the relevant term also denotes a 'large round fruit'). In addition, there are a number of languages in which the relevant term features a bound element conveying the meaning 'round thing;' these are Nuuchahnulth *hupk-imɬ* 'roundish.thing-CHUNK.SHAPED.OBJECT (which also means 'testicles,' as does Bislama *bol,* which goes back to Engl. *ball*, alongside 'sac' and "soft belly of coconut crab") and Cubeo *yajui-dɯ* 'game/sport-CLASS.ROUNDISH.OBJECT' (for this term, compare Cahuilla *qáwpiʔíl*, which is derived from

-qáwpi- 'to play a ball game,' such terms are also featured in Haida and Tuscarora, where it in addition colexifies 'setting sun') and Cayapa *sapuca* contains *puca* 'small round thing.'

However, various specific round things are colexified in many languages. In three languages of the Old World, Efik, Abzakh Adyghe, and Greek, 'ball' is colexified with 'bullet' (see Buck 1949: 907 on this extension in German; Adyghe also has terms colexifying 'ball' with 'cannon'). In both Hausa and Katcha, the relevant terms also denote the 'nut of the dum-palm' which is apparently used as a ball (in Hausa, also the Polo ball specifically and hence the Polo game as well as horse-racing inter alia). Bakueri *e̲we̲má* also means 'orange,' and Buli *gbeli* is also the name of "the wooden cylinder of the *buuri-* or calabash fruit that are used by sheperds as a 'ball,' when they play their game of 'hockey'." Ngambay *bbìri* also means 'parcel, container,' and *ékiyáy* also means 'placenta.' The Yoruba term *ìṣu* 'ball, lump of anything' (the association with 'lump' is shared with Samoan and Tetun, which has *ai-kabuar* 'tool-circle/lump' and many Indo-European languages according to Buck 1949: 907-908) consists of the verb *ṣu* 'to mould in a round form' and the prefixal nominalizer *ì-*. Kyaka k*onda* is also the term for a 'kind of yam;' the term denotes an oval-shaped ball similar in shape to the fruit of the *konda*-yam specifically. Badaga *seṇḍu* also means 'grass or straw ring,' Basque *bola* also 'skittle' and 'ball-bearing,' and Khalkha *bømbyge(n)* also 'bomb' and 'pellet used for crossbows.' Similarly, Welsh *pêl* also means 'pellet' and 'pill.' Cheyenne *htóhtséme ~ hohtsemo* also denotes a 'netted hoop,' and Lake Miwok *polóolo* colexifies 'ball' with 'dry oak ball' as well as 'baseball' and 'Ball Dance' specifically. Hawaiian *pōpō* also denotes a 'round mass,' 'cluster,' or 'bunch' generally, *pōkaʻa* (containing *kaʻa* 'to turn') also means 'coil,' 'roll,' and 'spool' inter alia. The basic meaning of Lenakel *nouanetp̶inami-* is 'urinary bladder,' which is reflected in the term's morphological structure: it is analyzable as /noua-netp-nami-/ 'fruit-belly-urine' (presumably inflated animal bladders are used as balls in games).

Otherwise, terms for 'ball' may contain verbs meaning 'to roll' or to 'roll around' (Efik, where the term also denotes the game played with the ball) or 'to kick,' as in Comanche *naʔsu̶hpeeʔ* 'small native-american football,' which is analyzable as /na-su̶h-peeʔ/ 'REFLX-kick-fall;' 'kick' is also the meaning Muna *sepa* 'ball plaited from rattan' assumes when used as a verb. By provenience contiguity, Rendille *imbíra* also denotes an object made of rubber or plastic generally, while Bora *máákiñiu* contains *máákiñ* 'rubber' and the classifier for small round objects *-iu*. In three languages of the New Guinea area, Buin, One, and Sko, there is a curious association between 'ball' and 'citrus': in Buin, 'ball' is colexified with a "kind of tree: a tree with spherical, lemon-like fruit about 10cm in diameter," One *malwa tala ~ maula rala* is analyzable as /malwa tala/ 'citrus round.thing,' and Sko *hangléúe*, which is a dvandva compound analyzable as /hang-lèue/ 'coconut-peanuts,' colexifies the meanings. Muna, spoken in the same broader region provides a hint towards an explanation: *golu lemo* is glossed as 'citrus fruit used as ball' (*golu* 'ball,' *lemo* 'citrus fruit'). Pawnee *rariickiriʔ* also denotes the 'baseball' and the associated game specifically which is also the case in Lake Miwok and Hawaiian, just like Buli *gbeli*, Carib *bal,* and Bwe Karen *bɔló* also denote the 'football' specifically. Ngaanyatjarra *purrpurl(pa)*, borrowed from English *football*, also denotes the game of the same name as well as "football carnival," (*tjaputjapu* likewise denotes both 'ball' and the game of football). Dongolese Nubian and Japanese

colexify 'ball' with 'bowl,' in Japanese due to phonological collapse of the English source words *ball* and *bowl*.

Other patterns include: Efik *obön* is also glossed as 'musketo,' Berik colexifies 'ball' with 'forefinger,' Mali *vaidebunggi* is used to refer to something of tennis-ball size and is also the name of a tree species, while Rotokas colexifies 'ball' with 'name' and 'letter.' Greek *mpálla* also means 'bale,' and Japanese *kyū* also 'nine.' A literal rendering of Carrier *nekhek* would be 'uses to be tossed about,' and Comache *naʔmahpeʔeʔ* would be translated literally as 'object thrown by hand.' Kiowa colexifies 'temple' and Santiago Mexquititlan Otomí *lobo* 'wolf' (due to phonological collapse of Span. *globo* and *lobo*). Central Yup'ik (Nunivak Island dialect) *yuguaq* is analyzable as /yug-(ng)uaq/ 'person-imitation,' and San Lucas Quiaviní Zapotec *pelo't* also means 'ball game.' Embera colexifies 'ball' with 'whole' and "surrounding district." Miskito *dans pulanka kum* is analyzable as 'dance play one.such.' Imbabura Quechua *muyu-ndin* is analyzable as 'seed-with,' whereas Hani *siilpuq* might contain *puq* meaning 'to swell' inter alia (*siil* is 'gold, yellow'). Rotuman *poro* also is the name of a bush inter alia, and the plural of Samoan *polo* also means 'roller bearings.' Finally, Vietnamese *bóng* also means 'shadow' inter alia.

74. The Bed

Representation: 88%
Motivated: 46.8%
Thereof Analyzable: 28.7% Thereof Colexifying: 17.7%
Thereof by Contiguity: 30.0% Thereof by Similarity: 10.0%
Recurrent associated meanings: sleep, furniture, lie/lie down, place, mattress, tree/wood, mattress, ground/floor, (bed-)room, quilt, thing, spread/spread out, nest, hammock, dream

Commonly, words for 'bed' contain verbs meaning 'to sleep' or 'to lie' (for which see Buck 1949: 480 in Indo-European), and they may be either of the derived or lexical type. Terms derived from 'to sleep' are found in Muna, Cahuilla, Chickasaw, Nez Perce, Central Yup'ik (Nunivak Island dialect, otherwise "sleeping bag, bedroom"), Cashinahua, Cavineña, Ancash Quechua, Fijian, Rotuman, and Samoan (here, there is a more complex term: *ulu-moe-ga* 'enter-sleep-NMLZ'). If terms are, however, of the lexical type, the second element may differ: it may simply be 'thing,' as in Mbum *fè-nâm* 'thing-sleep,' 'place,' as in Yoruba *ibùsùn*, which is analyzable as /ibi-isùn/ 'place-sleep,' as well as in Wichí (and note that Ngaanyatjarra, Abzakh Adyghe, Khalkha, and Fijian colexify 'bed' with 'place,' and Anggor inter alia with 'place,' 'chair' and 'floor'), or it may be the name of some other piece of furniture, as in Dadibi *pibo sai* 'sleep floor/table,' and also in Kyaka, Rao, Toaripi, and, questionably, Japanese (it is in fact relatively frequent for terms for 'bed' to also denote another piece of furniture or furniture in general, this pattern is found in Noni, Anggor, Baruya, Burarra, Mali, One, Rotokas, Sahu, Badaga, Nuuchahnulth, Tuscarora, Lengua, Rama, Hawaiian, Rotuman, and Bislama). Otherwise, complex terms involving 'tree' or 'wood' alongside 'sleep' are also found: Buli has *gaduok* /goa-duok/ 'sleep-wood,' Kyaka *luu palenge isa* 'sleep ?? tree/wood,' Copainalá Zoque *ʌŋguy* /ʌŋu-cuy/ 'sleep-wood/tree,' Kaingang *ka*

krẽ 'wood criciúma' (criciúma is a type of Bamboo; the complex term also denotes a 'field to beat beans'), and Maxakalí *mĩp-xap*, analyzable either as 'wood-stone/bead/seed' or 'wood-weave/knit/sew' (note also Yir Yoront *yo-way* 'high branches of a tree, stretcher, bed,' which is analyzable as 'tree-high.up' and Itzaj *tus'bil-che'* ~ *tus-che'* 'wood pile, rack, bed,' analyzable as 'stack(ed)-wood'). Furthermore, the Bororo term *boe enu pa* contains both terms for 'thing' (*boe*) and 'place' (*ba*) alongside a verb meaning 'to sleep' (*nu*). In Ngaanyatjarra, Aymara, Bora, and Hawaiian, '(to) sleep, sleeping' and 'bed' are directly colexified (in Hawaiian also other meanings are), in Aymara by the analyzable term *iki-ña* 'dream-INSTR.' In fact, Hawaiian colexifies 'bed' also with 'dream,' and relevant terms in Bora, Cashinahua, and Toba may or may not do the same, since the Spanish gloss 'sueño' is ambiguous between 'sleep' and 'dream.' Semianalyzable terms involving a verbs meanings 'sleep' or 'go to sleep' is found in Nez Perce, and one featuring a noun 'sleep' in Toba.

The other major association, that with verbs meaning 'to lie' or 'to lie down,' betrays the same mixture of derived and lexical terms. The derived variety is found in Efik (where the term also denotes a 'camp' and a 'rendezvous'), Carrier, Upper Chehalis, Ineseño Chumash, Kiliwa, Lakhota, Wintu (based on a verb meaning 'lie on ground' more specifically), Yuki, Jarawara, Fijian (where the derivation base is glossed as 'to lie in a place' more specifically, and hence the derived term also has the more general meaning 'location, place, position'), Malagasy, and Rotuman. For instance, Malagasy has *fandrìana*, analyzable as /fa-àndry-ana/ 'NMLZ-laying.down-NMLZ.' In Cheyenne, 'bed' is *šéešestôtse* /šéeše-hestôtse/ 'lie-thing,' in Haida it is *tay daan* 'lie place,' colexifiying 'basking place' and 'rookery.' In Tetun *tobafatin*, *fatin* is 'place' and *toba* can actually mean both 'to lie down' and 'to sleep,' and the same situation is encountered in Kwoma. Semianalyzable terms with 'lie'/ 'lie down' occur in Biloxi and Lengua (where the identifiable root also means 'to sit'). The Biloxi term in fact colexifies 'bed' with 'mattress,' and this pattern is also found in Buin, Abzakh Adyghe, Badaga, Upper Chehalis, Wintu, Tsafiki, and White Hmong, while Cubeo has *paraino*, with the root *parai-* also occurring in the term for 'mattress' and classifiers differentiating between the two referents. Similarly, Kiowa, Arabela, and Aymara colexify 'bed' and 'quilt.'

Another recurrent association realized by complex terms is that with the meaning 'to spread,' 'to spread out' (a pattern also evidenced in Old Norse and Church Slavonic, Buck 1949: 480): Swahili *kitanda* contains the verb *tanda* 'to spread, stretch' and a noun class prefix, while Rotokas *urua* is derived from *uru* "spread something out to sleep or lay [sic!] on," Ineseño Chumash *suwaskɨnɨmu'* is analyzable as /suwaskɨn-mu'/ 'to.spread.something.open-DERIV,' and Badaga *ha:sike* contains *ha:si* 'flat, to spread out' (a similar pattern may be discernible etymologically in Koyraboro Senni). In contrast, an association exclusively realized by colexification is that with '(bed-)room,' occurring in Ngaanyatjarra, Pawnee, Tuscarora and Guaraní, where the term in addition colexifies 'nest,' as is the case in Manange. Anggor, Baruya, Wintu, and Tsafiki colexify 'bed' with 'ground' or 'floor' (Wintu *łEh* also means 'spread,' 'hay,' and "hind legs of dead animals spread and dragging" and Lake Miwok *wéja* also 'base, bottom' inter alia; compare also Kashaya *cahti* 'bed' and *ca-* "be sitting on ground, floor"?), while Nuuchahnulth *čim?iɬ* is analyzable as /čim-'iiɬ/ 'right/ready-ON.THE.FLOOR/IN.THE.HOUSE,' and Wintu *pominpanas topi*

contains *pominpana* 'to lie on ground' (for the association with 'ground' in Baltic, see Buck 1949: 480). Furthermore, Cashinahua and Piro colexify 'bed' with 'hammock,' and Dongolese Nubian with 'bedclothing.'

Other associations include: Hausa *gado* inter alia also denotes the notions of 'inheritance,' 'bargain,' while Koyraboro Senni *daari* also denotes a 'bed-roll' and 'bedding.' Ngambay *tuwa* also denotes the 'straw to make a mat with' and other things, and Baruya *minye* can also refer to the "lower side of an ambush." Muna *koe* "wooden or iron bed" also means 'do not,' while Ngaanyatjarra *ngurra* can also mean 'camp' and 'living area.' Badaga *me:ḍe* also denotes 'cane, bamboo,' Chukchi *ajkol* also 'skins to sleep on,' Khalkha *oru* inter alia also 'vacant place, vacancy' and 'trace,' and Welsh *gwely* also 'family.' Lake Miwok *káama* also means 'crib,' and Nuuchahnulth *čim?iɬ* is also "Chimihl, the name for the passage between Congreve Island and the shoreline of Barkley Sound." Santiago Mexquititlan Otomí *xifi* also means 'to explain,' and Pawnee *kusaahkus* also means "be the site of a former dwelling, camp ..., be a campsite" and "be a field, playing field." This term is in turn based on the term *kusaar*, which is derived from *kus* 'to be sitting, be living,' and also denotes a 'seat' and 'place' generally as well as a station in a ceremony specifically. Tuscarora colexifies 'bed' with 'position,' 'space,' and 'stage,' and San Lucas Quiaviní Zapotec *ca'mm* also denotes an "iron cup filled with gunpowder and exploded like a firecracker." Central Yup'ik *aci ~ aciq* also means 'area below, area under,' Abipón *n-aoal-Ra* is analyzable as 'POSS.INDEF/3SG-bring-ABSTR,' and the Arabela term *maqueja* seems to be derived from *maque* 'footprint.' Cayapa colexifies 'bed' with 'mosquito net,' Chayahuita *pë'sara* is analyzable as /pë'sa'-ra/ 'palisade-CLASS.SMALL.THINGS.' Miskito *krikri* may also refer to 'bunk' and 'tapestry,' while Sáliba *jahaxoode* contains *jaha* 'feet.' Toba colexifies 'bed' with 'totora sedge raft,' and Hani *hhaoqzao* contains *hhaoq* 'pillow' and *zao*, acting inter alia as the classifier for beds. Kapingamarangi *hada* also means "platform of outrigger boom," while Bwe Karen *lo ʃɔ̂-mí= ʃɔ̂-a* is analyzable as 'stone rest-when=rest-eat.' Finally, Hawaiian *moe* inter alia also means 'to lie in wait, ambush' as well as 'to marry, sleep with,' Sedang *xóang* also means "to solve a problem, to divide," and Tetun *kama* also 'cradle.'

75. The Belt

Representation: 86%
Motivated: 36.4%
Thereof Analyzable: 26.8% Thereof Colexifying: 9.3%
Thereof by Contiguity: 22.9% Thereof by Similarity: 0.3%
Recurrent associated meanings: tie/bind/wrap/fasten, waist, strap, leather/skin, circle/loop, belly/stomach, bandage, rope, thing, middle, body

Most commonly, terms for 'belt' ('sash,' 'gird,' 'girdle,' 'waistband') are derived from verbs meaning 'to tie, bind, wrap, fasten,' or contain an element meaning 'waist,' or both (note that in Indo-European, 'belt'-terms are derived from the root **yōs-* 'gird,' which may have originally been **yōu-s-*, an extension of **yeu-* 'bind,' Buck 1949: 434). Pure derived terms, such as Efik *u-böp* 'NMLZ-tie' occur not only in this language, but also in Khoekhoe, Ineseño Chumash, Tuscarora, Central Yup'ik, Cashinahua (derived from a verb meaning 'tie around

waist' more specifically), Cubeo, and Toba. Muna and Hawaiian colexifiy nominal 'belt' with verbal 'tie (around waist),' and similarly, Bezhta has *tic'iyo*, a past participle of a verb meaning 'to put on,' Tuscarora *uyę?kwíhčreh* contains the verb *-yę?kwih-* 'to gird,' and semianalyzable terms featuring constituents meaning 'tie up' and 'tied' occur in Upper Chehalis, Copainalá Zoque, and Hawaiian. There are also complex terms with more than one lexical element, where one of them is a verb meaning 'to tie.' Kiowa has *tạn-pʻḤ'-ga* 'gap-be.tied-NOUN.POSTFIX,' Nuuchahnulth *łapẇanim* contains *łap* 'tie about' and *im* 'thing' (where *łap* in fact by itself means 'to gird, belt, tie about'), Bororo has *boe ekajejewu boe* 'thing' and *kajejewu* 'bind,' and similarly, Kaingang *vẽsogfín ja* is analyzable as 'bind.oneself thing.' Lengua has *aptete nipthlit*, with *tete* referring to "anything which is tied" and *ninthlit* to the 'waist,' Cubeo *jãrióicãva* is analyzable as /jãrióñ̶u̶-cãva/ 'tie-CLASS.CLOTH-RIBBON.LIKE' and *jãrió-me* as /jãrióñ̶u̶-me/ 'tie-CLASS.LIKE.THREAD,' and in Hawaiian, one of the words for 'belt' is *kama ʻaha* 'tie string.' More frequent, however, is the combination with 'waist' or 'stomach,' as in Abzakh Adyghe *bğə-rə-pχ(e)* 'waist-INSTR-tie,' which is also found in Mbum (where an additional element 'cord' is present), Sora, Upper Chehalis (where 'waist' is colexified with 'middle,' compare Fijian *i vau ni tolo-na* 'DERIV tie POSS middle.part-POSS'); similar denominations are Malagasy *fehikìbo*, analyzable as /fèhy-kìbo/ 'tying/knot-stomach' and Tetun *futu-kabun* 'bind-belly.' Jarawara has *makari tosi* 'cloth/clothing waist,' and furthermore, Aguaruna and Yaqui feature a monomorphemic term colexifying 'waist' and 'belt.' Derived terms with the lexical basis meaning 'waist,' such as Abipón *aat-Reki* 'waist-LOC,' are also found in Basque and Aguaruna. Other complex terms where one of the constituents means 'waist' or 'stomach' occur in Buli (*chiak gbain* 'waist leather'), Mbum (*sàl tàk* 'cord waist'), Kolyma Yukaghir (*aŋdil-amdi:* 'waist-bedding'), Yaqui (*toma jisumiam*, analyzable as /toma jisumia-im/ 'belly package[8]-PL'), Guaraní (*ku'a-sâ* 'waist-rope' and *ku'a-kua-ha* 'waist-have.holes-AGT'), Miskito (*maisa-wila* 'waist-ribbon'), Ancash Quechua (*tsiqlla watu* 'waist thread/strap'), and Vietnamese (*dây lưng* 'cord waist'). Semianalyzable terms with 'waist' occur in Bororo, Cayapa, Guaraní, Huambisa, Sáliba, and Wayampi, and the Ineseño Chumash term *tiwošokuš* contains the verb *tiwošok-* 'to wrap around the waist' and is literally translated as "something wrapped around the waist;" Copainalá Zoque has a semianalyzable term where the identifiable constituent means 'belly.'

Alongside the Buli term mentioned above, the meaning 'leather' or 'skin' also figures in Bora *ócáji-mɨ̃ɨ̃he* 'cow/tapir-CL.skin/leather,' Rama *bípuk* /bip-uk/ 'cow-skin' as well as in Efik, where 'leather' is colexified with many artifacts made from leather, among them 'belt' (the relevant term also has other meanings). Ngaanyatjarra, Lesser Antillean Creole French, Itzaj, Wintu, Cayapa, Macaguán, Sáliba, Rotuman, Samoan, and Bislama colexify 'belt' with 'strap' (see Buck 1949: 434 for evidence from Romanian), and three languages of the sample, Khalkha, Carib, and Hawaiian, colexify 'belt' with 'circle' and/or 'loop' (among other meanings by a number of terms in Hawaiian). Tetun *faixa* is also used with the meanings 'bandage,' 'lane on highway,' and 'track of record,' and Samoan *fusi* also denotes a 'bundle' or 'bandage' as well as "championship, final." Meyah *márféb efagá* is analyzable as 'cord body,' and Arabela *cajiniocuaque* as /cajinio-cuaqueya/ 'middle-body.'

[8] Original gloss is 'packege.'

Other associations include: Hausa *'damara* is also the name of a "geometrical figure consisting of two interlaced triangles" and, dialectally, "[a] coloured glass bangle." Noni *kecaw* appears to be derived from *caw* 'select/choose' by the noun class prefix *ke-*, Dongolese Nubian *gə̋s̋* also means to 'coquet, flirt, mince, simper, be spoilt' (Armbruster 1960: 131 relates the senses by way of the "underlying common notion" 'embrace'). Yoruba *ọ̀já* also denotes a 'head-tie,' Anggor *ŧitapurɨ* has the meaning 'bracelet' and might also be capable of referring to the 'belt,' and a literal translation of Dadibi *kibu wali* seems to be 'pig wind.' Muna *bhida* also means 'shroud.' Yir Yoront *maq* colexifies 'belt' with 'bottom, lowest part,' and *monporm* is the Yir Yoront term for both 'possum, fur of possum' as well as a belt made from possum fur. Japanese *obi* is a nominalization based on a verb meaning 'to wear,' and Welsh *gwregys* also may refer to a 'truss,' and *bad*, an obsolete term, also means 'plague.' Biloxi *a´xkido^{n}ni´* contains a verb meaning 'to wrap,' Cahuilla *tépaqal* is derived from a verb meaning 'to tighten a belt,' while in Comanche, *kohinehkiʔ ~ kohineekiʔ* has been extended to also denote a 'G-string.' Kashaya *phaʔɨ̇ʂaṭíʔ* appears to contain the instrumental prefix *pha-* 'by wrapping' and *ʔɨ̇ʂaṭ-* 'to hit hard and hurt.' Lake Miwok *mitúpponi* is analyzable as /míttu-pponi/ 'count-AGT.' Tuscarora *yeʔnhęhθuʔnarhúhsthaʔ* contains *-(i)ʔnhęhθ-* 'rawhide strip' and *-^{n}u'narhu-* 'hook.' Wappo *pá·haʔ* also means 'straw' (presumably due to the collapse of Span. *paja* 'straw' with *faja* 'strip, waistband'), wheras Wintu *lakum* contains *lak* "get caught, trip, hook; embrace, pinch," and *λa·q ċopċi* is used for rattlesnake skin that is worn as a belt specifically (and is indeed analyzable as 'rattlesnake skin'). Bororo *aie-wora* appears to be derived from *wora* 'leg' and also may refer to a 'Bororo tanga.' Lengua *yukma tama* is analyzable as 'skin.skirt string,' and Fijian *i oro* as 'DERIV clasp.' Hani *juqzaoq* appears to contain *juq* 'loose' (also meaning 'to throw') and *zaoq*, acting inter alia as a classifier for bundles. As a verb, Kapingamarangi *duu* means 'to stand' and 'to stop,' and Lenakel *katovɨt* is derived by the instrument nominalizer *k-* from the verb *atovɨt* "to put on clothing by wrapping it around self." Rotuman *fạli*, as a verb, also means "to thresh with a belt or strap." Bislama *strap* also denotes a 'seat belt' and 'fan belt' inter alia, and Yay features semianalyzable terms involving a constituent meaning 'man.'

6.2.2.76. The Boat

Representation: 70%
Motivated: 33.1%
Thereof Analyzable: 10.7% Thereof Colexifying: 23.3%
Thereof by Contiguity: 7.1% Thereof by Similarity: 6.9%
Recurrent associated meanings: canoe, water, trough, raft, vessel, collamon, airplane, vehicle, fire

Motivated terms for 'boat' (which is not distinguished here from 'ship') are relatively rare. Quite common are patterns of colexification that are somewhat similar to that of English *vessel*, that is, terms that denote both a receptable and container (for goods or other things) and a means of transportation. This situation is encountered in Basque and Central Yup'ik (and in Ancient Greek and Sanskrit, Buck 1949: 730); the Basque term also denotes a 'case,' 'sheath,' 'carton,' or 'pot,' and the Yup'ik term also a 'tray' specifically. Further-

more, Hausa, Mbum, Khalkha, Chickasaw, Pipil, and Tuscarora colexify 'boat' with 'trough.' The relevant Hausa term *jirgi* is also unique in being also applicable to a railway train (the two relevant Hausa terms have still further meanings). Further, Khoekhoe *ǂgaub* is also the name of an 'elongated bowl,' Gurindji *kartiyi* and and Yir Yoront *pinarr* are also used as the name of the 'coolamon,' an indigenous Australian carrying vessel, as well as, in Yir Yoront, for a "deep wooden oval dish," Badaga *teppa* also denotes a "temple pond" as well as an "artificical tank," Wappo *khéye* also denotes a 'cradle basket,' and Embera also a 'chest,' while Fijian *waqa* inter alia can also refer to a 'box, case, container' generally.

Complex terms for 'boat' are of a variety of structural types, and most often make reference in some way to 'water.' Kanuri has *mààrá njî-bè* 'vehicle water-of,' Yoruba *ọkọ̀-oju-omi* 'vehicle-eye-water' (alongside *ọkọ̀ kekere* 'vehicle little'), and 'boat' and 'vehicle' generally (as well as 'conveyance') are colexified in Quileute. Kaluli has *ho:n ko:su* 'water airplane' (this term denotes a modern-type boat introduced in the colonial era; note also that 'airplane' and 'boat' are colexified in Khalkha, which features a very general term for any kind of vehicle), Blackfoot *aahkioohsa'tsis* /yaahkioohsi-a'tsis/ 'travel.on.water/travel.by.boat-INSTR,' Cheyenne *amóehestsestôtse*, containing the prefix *am-* 'along by water' and *hestôtse* 'thing,' Chickasaw *okokaaittanowa'*, analyzable as /oka'-okaaittanohǫ́wa-'/ 'water-walk-NMLZ,' Comanche *pawobi,* analyzable as /paa-wob̲i/ 'water-board' (in the Kwahere dialect, there is an additional constituent meaning 'horse'), Santiago Mexquititlan Otomí *bojä dehe* 'iron/car water,' and Piro *gonu yapachro*, containing *gonu* 'water' and *ya* 'go.' There is a semianalyzable term in Guaraní. Moreover, there are two languages of South America with complex terms where one of the constituents is 'fire': Bora has *cúújúwa-mɨ* 'fire-SCM.transport and Wayampi *tata-l-ɛna* 'fire-of-place,' which colexifies 'fireplace.' In both cases, the terms denote a 'steamboat' specifically.

Santiago Mexquititlan Otomí has *motsa dehe* 'canoe water,' and generic terms for 'boat' are colexified with 'canoe' in Buli, Hausa, Mbum, Noni, Muna, Tasmanian (all varieties expect the North-Eastern one), Biloxi, Carrier, Cheyenne, Ineseño Chumash, Comanche, Haida, San Mateo del Mar Huave, Kiowa, Lakhota, Nez Perce, Tuscarora, Yaqui, Abipón, Guaraní, Sáliba, Toba, Wayampi, and Fijian, and with 'raft' in Buli, Dadibi, Badaga, and Yuki.

Other associations include: Hausa *komi* also denotes the "beds of an irrigated farm," while Dongolese Nubian *kúb* is also used with the meaning 'shuttle in weaving.' Lavukaleve *fela'koe* also means 'village.' Muna *kapala*, a Loanword from Bahasa Indonesia, also means 'leader, chief,' and indigenously also 'to sit on something raised,' while *bhangka* is also the name of a constellation and means "inner part of belly." Abzakh Adyghe *qᵒʲəḥe* contains *ḥe* 'to carry away.' Welsh *cwch* also means 'beehive.' Kiowa *kā'bout* contains *kā'-* 'to swim, to go by boat,' and Oneida *kahuweˑyáˑ* is also the term for the 'black ash.' Pawnee *rakuuhuuruʔ* is analyzable as /rak-huuhuur-uʔ/ 'tree/wood-floating-NOM.' Chayahuita *panca nansha marë pa'tërin-so'* is (semi-)analyzable as 'big ?? sea leave-3SG.SUB.' Toba *lllicta ~ lllocta ~ lllicota* is derived from *illigot ~ illogot* 'rows,' while Wayampi *ɨa* is also the name of an ant species. Fijian *velovelo* is also the name of a string figure, Hani *loq* also means 'to rinse a container with water' and acts inter alia as a classifier for irrigated fields, Bwe Karen *khlí* is

also the name of a month, Hawaiian *moku* also means ‘island’ inter alia, while Kapingamarangi *waga baalii* (containing *waga* ‘canoe’) also means ‘grasshopper.’

77. The Car

Representation: 66%
Motivated: 41.9%
Thereof Analyzable: 27.8% Thereof Colexifying: 15.2%
Thereof by Contiguity: 18.5% Thereof by Similarity: 10.2%
Recurrent associated meanings: vehicle, cart/carriage, machine, roll, self, run, thing, train, motor vehicle, canoe, sled, twist, fire, house, ride, land

A very interesting pattern in terms for ‘car’ is that some of them contain elements meaning ‘move’ and ‘self’ (which is in some of the relevant languages conveyed by a reflexive marker). This is ultimately the literal meaning of the Graeco-Roman hybrid compound *automobile*, the constituents of which have just this semantics. This denomination also recurs in other languages alongside Greek, where it is (at best) semianalyzable in the modern langage, for instance in Carrier *tiqerh-nekhĕs* ‘proceed-by.itself,’ and also in Nez Perce, while Kashaya has the somewhat similar *yu?dul q^{h}ayam?*, analyzable as /yu?dul ^{h}q^{h}ay-am-?/, containing *yu?dul* ‘self’ and $^{h}q^{h}ay$ ‘run,’ and the literal translation of Lakhota *iyéčhįkįyąke*, according to the consulted source, is “it runs by itself” (for the element ‘run,’ compare also Ngambay *né kàyn ngɔru* ‘thing run fast,’ Samoan *ta‘a-vale* ‘run.freely-bad/of.no.use/ordinary,’ and Kiliwa *(?+)wa?=kw-s-?+hin* ‘(DN+)house=WH-IRR-DN+run’). Tehuelche has the similar term *wawere:nk* /waw-ʔere:-n-k/ ‘sole-walk-NMLZ-MASC,’ which can also refer to a ‘bachelor’ (there are further variants of this term). The presence of this pattern in many languages of the Americas almost suggests calquing, although, obviously, this would presuppose knowledge of the meaning of the constituents. Whether indeed these terms were calqued under European influence (perhaps mediated by missionaries?), coined independently, or a mixture of both remains an open question. Further, somewhat similar is the Fijian term *qiqi toso* ‘roll/vehicle move.of.itself,’ and a term containing a verb meaning ‘to roll’ is also featured in Chickasaw (*itti' chanaa palhki'* ‘wood roll be.fast’), while the association with ‘rolling’ is by direct colexification in Hawaiian (among other meanings); compare also the Ngambay term with a constituent ‘speed’ mentioned above as well as Cheyenne *ameohe-hestôtse* ‘go.by.quickly-thing,’ Central Yup’ik *akag-cuun* ‘roll-device.for’ (this term colexifies ‘wheel’ and ‘axle;’ there are other dialectal variants), and Hupda *pəpód-teg* ‘roll-thing.’

Cheyenne also has the alternative term *am-âho'-hestôtse* ‘along-by.heat-thing,’ and, somewhat similarly, there are two languages in the sample, Bora and Wichí, in which relevant terms feature an element meaning ‘fire’ (*cúújúwa-mɨ* ‘fire-SCM.transport’ and *wej itoj* ‘end fire’ respectively). In two sampled languages, terms for ‘car’ make reference to the loud noise it produces: Mali has *araun'ga mētki*, analyzable as /araun-ka mēt=ki/ ‘sound-M.SG in=3F.SG’ (this term may also refer to a motorcycle), and Acoma *tárárágá*, analyzable as /tárərəka-'/ ‘be.roaring-INSTR.’ In Efik and Kiliwa, there is a metaphorical transfer from ‘house’ to ‘car:’ the relevant terms in these languages are *ufök enañ makara* ‘house cow

European' and *(?+)wa?=kw-s-?+hin* '(DN+)house=WH-IRR-DN+run,' as already mentioned. Tuscarora has *u?θréhčreh*, containing the root *-(i)?θ(e)*r- 'to ride,' and a similar term, containing the root *-?sle-* 'drag, ride, drive, trick someone' is found in the related language Oneida. Central Yup'ik *nuna-kuar-cuun* is analyzable as 'land-over-device.associated.with,' and Piro *tš̲ix̲iyapatš̲ro* contains *tš̲ix̲i* 'earth, land' and *ya* 'to go.' Guaraní *mba'e-jere* is analyzable as 'thing-twist' (there is a further semianalyzable term with *mba'e*), while 'to twist' is inter alia colexified with 'car' in Hawaiian.

There are also associations with other vehicles: Swahili, Abzakh Adyghe, Basque, Japanese, Khalkha, Bora, Kaingang, Malagasy, and Samoan colexify 'car' with 'cart' or 'carriage' (Bora by the analyzable term *íjchí-e-mɨ* 'come.ashore-belong.to-SCM.transport;' it cannot be excluded that there is an error in the source), Comanche, Pawnee and Wintu with 'train' (Pawnee and Wintu have borrowed English *car* or the plural form *cars*), Upper Chehalis colexifies 'canoe' (while Kwoma has *gaba veyi* 'whiteman/ghost canoe'), and Kildin Saami, Ineseño Chumash, Kashaya, Lake Miwok, and Wappo colexify 'car' specifically with 'machine' generally (all have borrowed the respective terms from contact languages which in turn ultimately go back to Latin *machina*; Kashaya also has optional complex terms on the basis of this term). Moreover, in two sampled languages spoken at high latitude, an association with 'sled' is found, in Central Yup'ik by colexification, and in Ket by the analyzable term *ēɣ suul* 'iron sled.' Moreover, Koyraboro Senni, Rendille, Swahili, Japanese, Ket, Khalkha, Oneida, Tuscarora, Hupda, Lenakel, White Hmong, and Bislama colexify 'car' with '(wheeled) vehicle' in general, and Koyraboro Senni and Ngaanyatjarra with 'motor vehicle' (otherwise, the presentation does not differentiate between 'car,' 'truck,' 'bus' etc.). Similarly, Koyraboro Senni *moobil-ize* is analyzable as 'vehicle-child,' Yoruba *ọkọ̀ ayọ́-kẹ́lẹ́* as 'vehicle fanciful-manner,' and Fijian *qiqi toso*, as already mentioned above, as 'roll/vehicle move.of.itself.'

Other associations include: Buli *logri* also means 'to give way,' 'to avoid,' and other things, while Kyaka *karo* also means 'dirt' and 'grime' inter alia (the meaning 'car' is due to borrowing from Tok Pisin). Muna *mintoro*, an obsolete term, is related to *ntoro* 'to turn, rotate.' Ngaanyatjarra *yurltu* also inter alia means 'empty' and 'hollow tree,' the common denominator of the meanings probably being that a car is "hollow" in the sense that it provides space for sitting in. Sahu *'oto* also means 'to cut,' Basque *auto* also 'edict, judicial decree' and 'mystery play, religious play,' and Khalkha colexifies 'rook in chess' inter alia. Welsh *car* also means 'trap.' Blackfoot *iitáísapópao'p* contains *sap* 'inside' and *opii* 'to sit.' Upper Chehalis *xʷiyúyəqs* is derived from *xʷiy-* 'to cut off;' the term might be a loan translation from Chinook Jargon. Comanche *naʔbukuwàaʔ* is analyzable as /na-puku-waa/ 'REFLX-horse-horn.sound,' while Pawnee has *kiriiraawis* /kiriir-raawis/ 'anus-smoke' and variants of this term. Santiago Mexquititlan Otomí colexifies 'car' with 'tool' and 'iron.' Wintu *pʰuλurumes*, a term found in the Trinity County dialect, contains *pʰuλ* 'to blow' (literal translation provided in the source is "puffing one"). Yuki *lulą̀mą̀l* is possibly literally 'oil puller.' Mandarin colexifies 'car' with 'chariot,' White Hmong *hov* also means 'short' (in the sense of 'vehicle' it is a borrowing from Lao), and Bislama *trak* (presumably due to collapse of Engl. *truck* with *track*) also means "footprint, spoor, track" and is also the name of a wheeled children's toy.

78. The Chair

Representation: 84%
Motivated: 53.0%
Thereof Analyzable: 40.6% Thereof Colexifying: 12.8%
Thereof by Contiguity: 38.5% Thereof by Similarity: 10.3%
Recurrent associated meanings: sit/sit down, furniture, wood/tree, place, thing, buttocks/bottom, saddle, throne, situation, floor, dwell

Clearly, the most frequent association as realized by morphologically complex terms is that with verbs meaning 'to sit' or 'to sit down' (as in Indo-European, Buck 1949: 482). The association is by derivation in Efik (*i-tie* 'NMLZ-sit;' this term inter alia also means 'situation,' which is also colexified in Kapingamarangi, or 'state' figuratively), Burarra (where 'to sit on' is colexified with 'put one's weight on, tread on, step on'), Gurindji, Mali (colexifying also 'meeting' and "sitting of parliament"), Chukchi, Sora, Blackfoot, Cahuilla, Upper Chehalis, Chickasaw (where the relevant term colexifies 'toilet'), Ineseño Chumash, Kashaya, Nez Perce, Nuuchahnulth, Yuki, Central Yup'ik, Abipón, Bora, Carib, Cashinahua, Guaraní, Miskito, Imbabura Quechua, Tehuelche, Toba, Yanomámi, Fijian, and Samoan, while Hawaiian directly colexifies verbal 'to sit' with 'chair' inter alia. When terms contain a second lexical element, this is often 'wood' or 'tree,' as in Baruya *namwaalyɨta* /namwaalɨmo-yɨta/ 'for-sitting-wood.' This pattern is also found in Ngambay, Kyaka (where 'sitting' is colexified with 'living'), Yir Yoront (here the term contains more constituents: *yo-penpn pam nhin+nh* 'wood-flat person/body sit.down+REL'), Ket, Kiowa, Lakhota, and Copainalá Zoque (other terms involving a constituent meaning 'wood,' but not 'sit,' are Swahili's *kiti*, consisting of a noun class prefix and *mti* 'tree' and Itzaj's *k'an-che',* which is perhaps analyzable as 'support-wood.' Note also the similarity between Yoruba *àga* 'chair' and *agà* 'tree'). Otherwise, 'thing' figures as the second constituent alongside 'sit' or 'sit down,' as in Katcha *nimo ma th-andanɛ* 'thing GEN ??-sit,' also in Mbum, Ngambay, Dadibi, and Bororo ('thing' is also the meaning of the identifiable constituent of the Lenakel word for 'chair'). In still other languages, 'buttocks,' 'bottom' or the like is attested as the meanings of the second constituents, as in Cheyenne *táxe'êséestôtse*, which is analyzable as /táxe-'esé-e-hestôtse/ 'upon-buttocks-sit-thing.' Similar terms are found in Pawnee and Bwe Karen (in addition, Toaripi has *kiri posa* 'buttocks platform,' and Tuscarora *uthečhráhkweh* contains roots meaning 'buttocks' and 'to collect'). Furthermore, Khoekhoe has *ǂnû-aiǃnao-s* 'sit.down-front.bench-3SG.FEM,' Wintu *kenλaˑ-s-poˑm* 'sit-??-land/ground/floor' (compare colexification of 'chair' and 'floor' in Anggor), Piro *tuplapiye* contains *tuplata* 'sit down, be seated' and *pi* 'rod,' Bwe Karen has *lo ʃɛ́ná* 'stone sit,' and White Hmong *roojzaum* contains *rooj* 'article of furniture' and *zaum* 'sit.' Semianalyzable terms with 'sit' or 'sit down' are furthermore featured in Upper Chehalis, Central Yup'ik, and Lengua (where 'sit' and 'lie' are colexified).

In Embera, Hawaiian, and Samoan (in both Polynesian languages, 'to sit' is colexified with 'to dwell' inter alia), 'chair' is colexified with 'saddle' among other meanings (given the morphological structure, this also seems to be the meaning of Miskito *aras nila pila* 'horse back down'), while in Noni, Anggor, Basque, Khalkha, Nivkh, Nez Perce, the

Santo Domingo de Guzmán dialect of Pipil, Quileute, Tuscarora, Wappo, Aguaruna, Aymara, Bororo, Cashinahua, Embera, Huambisa, Hupda, Kaingang, Lengua, Piro, and Yanomámi, terms meaning 'chair' may also refer to any piece of furniture in general or denote another specific piece of furniture (other than 'stool,' which is disregarded here), such as a 'bench' or a 'couch.' Words for 'chair' also have the general meaning of 'place (of something)' in Efik, Anggor, Ineseño Chumash, and Bororo. Burarra and Aguaruna colexify 'chair' with 'throne' (compare the origin of 'throne' from an Ancient Greek word for 'chair' more generally, Buck 1949: 481).

Other associations include: Buli *zukpaglik* contains *zuk* 'head' and *kpagli* 'to rest one's head.' The original meaning of the term was 'headrest' rather than 'chair,' a situation which is mirrored by synchronic colexification in Rendille. Mali *achut ngēthathengbēt* is derived from a verb meaning 'to lean back.' Sko has *fú-jéng* 'post.of.house/corner-place' (though also note *fú* "bottom of a four-legged animal"). Bezhta *q'ō* also means 'anvil,' and Welsh *cadair* also 'cradle' and 'udder.' The Biloxi term *ya´xoxoⁿni´* contains *xoxo* 'to swing.' Oneida *anitskwahlákhwaʔ* is analyzable as /an-itskw-hl-hkw-waʔ/ 'SRFLX-seat/part.of.body.one.sits.on-set.on.top.of/place.on-INSTR-??.' Tuscarora *uʔθkwéhseh* also denotes a "cutting block" as well as a "round block of wood, piece of a log." Yaqui *banko* also denotes the 'bank,' and Aguaruna *ekeémtai* contains *ekeémi* 'set on top' and the instrument nominalizer *-tai*. Guaraní *apyka* contains *apy* 'extreme point, deposit,' and Huambisa *ekemtai* possibly *eken* 'room, bedroom.' Ancash Quechua *silla* (< Span. *silla*) may also refer to a 'frame' or 'harness' (original Spanish gloss is 'montura'), Wichí *to-wej-w'et* is analyzable as 'POSS.INDEF-queue-place,' and Yanomámi colexifies 'desk, console.' Kapingamarangi *lohongo* also means 'situation, status' and 'office.' White Hmong *tog* also means 'block,' and Rotuman *nofoʻa* also 'chief.' Sedang *táng* also means 'to look for,' White Hmong *tog* also 'to sink' and "half-way point," Bislama *jea* (< Engl. *chair*) also means 'ticket, seat in a plane' as well as 'to eat.'

79. The Clock

Representation: 70%
Motivated: 60%
Thereof Analyzable: 27.6% Thereof Colexifying: 32.4%
Thereof by Contiguity: 51.4% Thereof by Similarity: 6.1%
Recurrent associated meanings: sun, hour, day, time, measure, bell, season/age, eye, imitate/imitation, watch, strike

There are a wealth of terms of different types for the 'clock' (or 'watch,' which is accepted as a proxy for this concept) that make reference in some way to the 'sun' (compare Latin *sōlārium*, derived from *sōl* 'sun,' Buck 1949: 1002). Burarra, Gurindji, Toaripi, Upper Chehalis (with a vowel change), Cheyenne, Comanche, Kashaya, Kiowa, Wappo, Wintu, Bora, Cubeo, Jarawara, Yanomámi, and Kapingamarangi directly colexify 'sun' with 'clock' (in some languages, as discussed in section 60, also with 'moon' and sometimes also 'month,' in Burarra, as discussed in section 57 and § 6.2.2.1., also with 'star' among other meanings, in Kiowa also with 'summer,' in Wappo also with 'calendar,' and in Jarawara also with

'thunder' and 'lightning,' see Dixon 2004: 71 for the history of this association). In New Guinea, associations with 'eye' are attested in Kaluli (*of-a:-si* 'sun-GEN-eye') and Meyah (*mówa eitéij* 'sun eye'). In two languages of South America, Aguaruna and Wayampi, the words contain constituents meaning 'sun' and 'to imitate' or 'imitation,' while in Yuki and Cashinahua (where 'sun' is colexified with 'hour,' for other associations with 'hour' see below), the second element is a verb meaning 'to watch.'

There are also other complex terms of the lexical type where one constituent is 'sun.' In Wichí, *tokafwala* is derived from *fwala* 'sun, day,' Baruya *nyihaanya* is literally 'sun-go,' Rotokas has *ravireo vetaveta-pa* 'sun count-DERIV,' Biloxi has *ina´do^{n}hi´~ ina´do^{n}´ho^{n}ni´*, containing *ina´* 'sun' and *do^{n}* 'to look at, see,' in Carrier, *sa-dzî* 'sun-heart' is 'watch' and *sa-dzî-tco* 'sun-heart-AUG' is 'clock,' Chicakasaw has *hashi' kanalli isht ithana-'* 'sun/moon move.to.a.new.location with know-NMLZ,' Pawnee *sakuhka?iirus* is analyzable as /sakuur-kaiir-hus/ 'be.a.day/sun-put.in/on-IPFV,' Arabela has *pananu shanacutaja* 'sun measure,' and Huambisa *etsa nakumkamu* 'sun drawing;' Yanomámi has a redundant complex term involving *mothoka ~ motoka* 'sun, clock' and *mɨɨ* 'see,' the literal translation of which is "that which serves to see the sun." The notion of measuring time, as in Arabela, also plays a role in the conceptualization of 'clock' in a number of other languages. In Japanese, 'time' is the other meaning figuring in complex terms of the lexical type: *to-kei* is analyzable as 'time-measure.' This is the only language with this particular configuration, though there are others which betray an association with 'time,' as is discussed further below. When 'day' rather than 'sun' is the meaning of the contiguous constituent, it is more frequent to have terms with the meaning 'count, measure' as the second constituent, as in Hupda *wág tæ̃?kéy* 'day measure.NMLZ' (thus precisely parallel to Old English *dæg-mǣl*, Buck 1949: 1003), and also in Upper Chehalis, Lake Miwok, Pawnee, and Quileute, where an additional element meaning 'thing' is present. Furthermore, Kiliwa has *maat=kw-p-c-?+wir-u?*, which is analyzable as 'REFLX=WH-MP-INST/MOUTH-DN+measure-OBL,' and an easier palpable literal translation offered by the lexicographer is "it measures itself." Other terms in which 'day' figures are Guaraní *arairû ~ arirû* /ára-irû/ 'time/day/sky-companion' and Malagasy *famantaranàndro*, analyzable as /fa-fànatra-ana-àndro/ 'NMLZ-known-NMLZ-day.' In Toaripi and Comanche, 'clock' is colexified with 'day' (and also with 'sun,' compare section 60), and Blackfoot has a term derived from a verb meaning 'be day.' Yoruba, Burarra, Khalkha, Hani, Samoan, Bislama, and Sedang colexify 'clock' with 'time' or 'time of day' specifically (Muna *dhamu* also means 'be time for something' when used as a verb as well as "tonic made of medicinal herbs") and an analyzable terms with 'time' is featured, alongside Japanese as already discussed above, in Piro (*hohi himata-tš̲aro* 'time know-??'); moreover, there are semianalyzable terms in Khoekhoe and Haida, with the other element diachronically related to a verb meaning 'to see.'

Frequently, 'clock' is colexified with 'hour' (see Buck 1949: 1002 for Indo-European evidence, e.g. from Middle High German). This occurs in Buli, Hausa (also with 'good luck' inter alia), Dongolese Nubian, Swahili, Berik, Muna, Abzakh Adyghe, Bezhta, Khalkha, Laz, Kildin Saami, Santiago Mexquitlan Otomí, Central Yup'ik, Chayahuita, Tehuelche, Hani, Bwe Karen, Samoan, and Sedang (colexifying also other meanings). In addition, Basque has *ordu-lari* 'hour-AGT,' Ket *časaŋ* 'watch' is the plural form of *čas* 'hour'

(as with Russian *časy*), Oneida *kahwista?ékta?* is analyzable as /ka-hwista?ek-ht-ha?/ 'NEUT.AGENT-be.the.time.or.hour-CAUS-HAB.' The Hausa and Khalkha terms for 'clock' have very broad semantic latitude, and may refer also to temporal concepts such as 'season' and 'age.' In four languages of the Old World, Yoruba, Khalkha, Sora, and Mandarin, 'clock' is colexified with 'bell' (common in Celtic and from this source, Germanic, but also Latvian, Buck 1949: 1003-1004); Sora also colexifies 'gong.' In Efik, the word 'clock' contains *mi'a* 'to strike' and reflexive markers, and similarly, in Tuscarora, *kawenę́·t?ehs* contains *-wenę́T-* 'iron' and *-(i)?e(k)-* 'to strike.'

Other associations are: Buli *bang* also means 'bracelet, wristlet,' and Hausa *sa'a* inter alia also 'good luck' and "a propitious time." Nez Perce *liklí·ńes* is analyzable as /likilí·-ńes/ 'go.around-INSTR.' Toba *lhuaxashi* contains *hua* 'forearm, hand,' Tehuelche colexifies 'clock' with 'alarm clock,' Fijian *kaloko* also denotes a 'very large kava, or roll of sinnet' (the meaning 'clock, watch' is due to borrowing from English). Kapingamarangi *laa* also means 'sail,' Mandarin *biao3* is elliptical for *shou3-biao3* 'hand-meter,' and is hence also used for other measuring instruments. Samoan *uati*, as a verb, also means 'to watch someone,' and Bislama *klok*, rarely, also may refer to the 'flamboyant, flame tree,' "because its leaves close up at night."

80. The Glasses

Representation: 58%
Motivated: 62.4%
Thereof Analyzable: 55.8% Thereof Colexifying: 6.6%
Thereof by Contiguity: 31.0% Thereof by Similarity: 24.6%
Recurrent associated meanings: eye, glass, mirror, see/look, metal, cover, imitation, put/set

Motivated terms for 'glasses' are, perhaps unsurprisingly, clearly more often analyzable than colexifying, and, even less surprisingly, these terms with very few exceptions have one constituent meaning 'eye' by contiguity. As for the second constituent 'glass' is common, as in Muna *mata tonde* 'eye glass.' Such terms are also found in Efik, Yoruba (where an additional constituent meaning 'vision' is present), Kyaka, Ngaanyatjarra, Yir Yoront, Cubeo (additionally suffixed with the classifier *-rʉ* for roundish threedimensional objects), Miskito, Hawaiian, and Yay. Furthermore, Hausa, Sedang, and Vietnamese directly colexify 'glasses' with 'glass,' Welsh has *gwydr-au* 'glass-PL,' and White Hmong *tsom-iav* 'look-glass.' Also common are terms based on a metaphorical comparison with a particular object made of glass, namely the 'mirror.' This is the case in Yoruba, Kyaka, Ngaanyatjarra, Hawaiian, Yay, so that these languages fall both in the categories exhibiting an association with 'glass' as well as 'mirror.' Languages in which complex terms for 'glasses' are found involving constituents meaning 'eye' and 'mirror,' without colexification of 'glass' and 'mirror,' are Buli, Koyraboro Senni, Mbum, Abzakh Adyghe, Badaga (where the 'mirror'-word also means 'lens'), Santiago Mexquititlan Otomí, Mandarin, and Rotuman. In the latter language, for instance, 'glasses' are called *maf tiro* 'eye mirror.' Further, Hausa, Bislama, and White Hmong directly colexifies 'glasses' and 'mirror' (and Hausa, as noted

above, also colexifies the substance 'glass,' while the Bislama term colexifies 'glasses' and 'mirror' and may also refer to other objects entirely or partially made of glass). Things get somewhat more complicated when one notes that some languages, in particular ones in which 'glass' is an item of acculturation, colexify 'glass' with 'metal.' Yir Yoront and Cubeo are of this type, and therefore, relevant complex terms for 'glasses' also betray a lexico-semantic association with 'metal.' There are, however, also complex terms for 'glasses' with this pattern where 'glass' and 'metal' are not colexified. These are Japanese, Cheyenne, and Nez Perce, which, for instance, has *kicú·ynim sílu* /kicúy-nim sílu/ 'metal-POSS eye;' in Cheyenne, there is an additional constituent meaning 'thing.' In Acoma and Central Yup'ik, glasses are conceived of as an 'imitation' of the eye (Yup'ik has a dedicated postbase with this meaning). Here, 'glasses' are called *ʔúwána'aẓáṅi*, underlyingly /húwana'ani-ẓaṅi/ 'eye-imitation' and *iinguak ~ iiguak ~ iigguak* /ii-(ng)uaq/ 'eye-imitation' respectively. In two languages of the Americas, terms make reference to the fact that one 'puts' or 'sets' glasses onto the eye: Kashaya *huʔuy duteʔtiʔ* contains *huʔuy* 'eye' and *duteʔ-* 'to put,' and in Kiliwa, 'glasses' are *ny-ʔ+yuw-l=t+papu-uʔ* 'POSS-DN+eye-ILL=OBJ+set-OBL,' literally, according to the lexicographer, "things one sets on one's eyes." Terms which directly make reference to the fact that glasses are used to aid seeing are surprisingly few. Chickasaw has *ishkin ishpisa'* ~ *ishkinshpisa'* which is analyzable as /ishkin isht pisa-'/ 'eye with see-NMLZ;' a similar term is only found in Upper Chehalis and Wintu (where there is a further term containing an element 'eye' and 'to catch fish in a net, hold out a net to catch fish'). Hausa features a derived term from a verb meaning 'to look,' Fijian has *mata-ilo-ilo* 'eye-look.at.reflection-RED,' White Hmong, as already mentioned, has *tsom-iav* 'look-glass,' and Rotokas *osireipava sisiro* appears to contain *osireito* 'eye' and *sisiro* 'inspect, stare, look intently.' Pawnee, Kaingang, and Yanomámi feature terms for 'glasses' involving constituents meaning 'eye' and 'to cover,' Pawnee, for instance, has *kiriktahkuukuʔuʔ*, analyzable as /kirik-raarkuuku-uʔ/ 'eye-covering-NOM.' A very similar term is found in Yanomámi, and Kaingang has *kanẽ kri táv* 'eye above cover.' Other complex terms with one constituent meaning 'eye' are Noni *ɛ-jisɛ bala* '6-eye foreign,' Anggor *hoe hɨmboarɨ* 'water eye,' Baruya *kwaari'matɨnna* /kwaari'mata-tɨnna/ 'plastic-eye,' Ngaanyatjarra *kurungkatja*, which contains *kuru* 'eye' and *katja* 'son' and also means 'eye ointment,' Bezhta *häydä*, which is grammatically the plural of *häy* 'eye' (and hence, can also refer to the 'eyes'), Nivkh *njaχ-ajs* 'eye-gold' (the term is used to refer to a small piece of metal which is put on the eye of the deceased, and has presumably been extended to cover 'glasses' later from there on), Kolyma Yukaghir *šöjd-aŋd'ə* 'stone-eyes' and *aŋd'əd-ajbi:* 'eye-shadow,' Comanche *pui tsaʔnikaʔ* 'eye underwear' (but compare *tsaʔatsitɨ* 'to inspect'?), Bora *hálluɟɨ* /hálluu-jɨ/ 'eye-CL.disc,' Guaraní *tesa-joa(py)* 'eye-together' and *tesa-irû* 'eye-companion,' Hupda *kəwəg-túʔ* 'eye-immerse.NMLZ,' Piro *sutsa-yhalu* 'clutch/pinch/grip-eye,' Wayampi *ɛapalitõ* /ɛa-palitu/ 'eye-sparkling,' Wichí *tot-telhu-hi-s*, containing *telhu* 'eyes' and the locative suffix *-hi* 'in,' and Malagasy *solomàso*, analyzable as /sòlo-màso/ 'substitute-eye.' Finally, Upper Chehalis, Tuscarora, Toba, and Samoan directly colexify 'eye' with 'glasses,' and semianalyzable terms are found in Kemtuik, Cahuilla, Piro, Yanomámi, and Hani.

Other associations are few: Hausa *madubi* is also used to express the affections of parents to their child as well as to refer to a sorcerer. Ngaanyatjarra *winta* (< Engl *window*)

also means 'window.' Carrier *naḳětṣełya* is the plural form of *naḳětṣeł* 'monocle,' and Tetun *ókulu* also means 'binoculars' and 'telescope.'

81. The House

Representation: 97%
Motivated: 29.7%
Thereof Analyzable: 3.0% Thereof Colexifying: 26.4%
Thereof by Contiguity: 13.0% Thereof by Similarity: 7.7%
Recurrent associated meanings: village, nest, room, family/lineage, roof, household, shelter, place, receptable, company, canvas cover, tepee, tepee cover, post, land

The most frequent association for 'house' (or 'building' generally, often also denoting 'home,' an association which is ignored here, just like cases when terms denote a specific type of house) is, by configurational contiguity, that with 'village' or more generally a group of houses, occurring exclusively by colexification in Katcha, Ngambay (which colexifies also 'country'), Kwoma, Mali, Ngaanyatjarra, Yir Yoront, Badaga (colexifying "isolated settlement" more precisely, as well as 'young'), Nuuchahnulth, the Cuisnahuat dialect of Pipil, Cashinahua, Jarawara, Miskito, Wayampi, and Bwe Karen (compare the cognacy of Ancient Greek *οἶκος, οἰκία* 'house' with Latin *vīcus* 'group of houses, village' and other evidence from Indo-European reported in Buck 1949: 458). By meronymy, some languages use the same word for 'house' and 'room.' These are Hausa, Badaga, Bezhta, Comanche, San Lucas Quiaviní Zapotec, Aymara, and Bwe Karen. Also by meronymy, Burarra, Arabela, Cubeo, Jarawara, and Tsafiki colexify 'house' with 'roof' (this pattern is common in a goup of Indo-European terms, Buck 1949: 458; Burarra also colexifies 'lid'). Presumably, this is also the motivation for colexification of 'house' with 'post' in Yanomámi (compare also Miskito *playa bila* 'post space,' which quite literally also denotes the 'space between two posts'). By functional similarity, Hausa, Khoekhoe, Kyaka, Muna, Badaga, Wintu, and Lenakel colexify 'house' with 'nest' (Muna also with 'web,' the Khoekhoe term *oms* is derived from the verb *om-* 'to build, construct,' compare Proto-Indo-European **domo-* ~ **domu*, from **dem* 'build,' Buck 1949: 458). Nunggubuyu, Waris, Basque, and Haida colexify 'shelter' (Nunggubuyu "stringybark shelter or dwelling" specifically), and Pawnee and Tehuelche 'canvas cover.' Ket *iŋɕus* contains *qu ǯ* 'tent,' Wintu colexifies these meanings (compare Slavic evidence reported in Buck 1949: 459), and Nez Perce *cóqoy* also means 'teepee top, smoke hole' ('teepee cover' is also the meaning of the relevant Pawnee term, and Comanche and Kiowa colexifies 'house' with 'teepee' directly; compare the cognacy of some Indo-European terms for 'house' with 'hide,' which go back to a root meaning 'to cover,' Buck 1949: 458). A pattern apparently particularly common in the Old World is metaphorical extension to 'family' or 'lineage,' found in the sample in Buli, Rendille, Abzakh Adyghe, Badaga, Basque, and Wintu, while Nuuchahnulth *maʔas* similarly also means 'tribe' (compare further the cognacy of the Ancient Greek and Latin terms mentioned above with Old Persian *viθ-* 'royal court, palace, family,' Buck 1949: 458). Similarly, Rendille, Kwoma, Badaga, and Lesser Antillean Creole French colexify 'house' with 'household,' and Central Yup'ik *enaᵉ,* Bororo *eda ~ jeta,* Jarawara *tabori/taboro,* Miskito *watla*,

and Wayampi *ɛna* are also used with the general meaning 'place' (the Jarawara and Wayampi terms also mean 'land'), and Badaga *mane* also with the meaning 'ground.' Basque and Greek colexify 'house' with 'company, firm' (the Haida term is glossed as 'establishment,' but it is unclear whether this actually refers to an enterprise). Lakhota *thí* also means 'to live, dwell' in verbal usage, while in Wintu, *bo·s* is derived from *bOh* "live, reside, remain, keep; stay, be in a sitting position, sit, dwell, stay" by means of the generic aspect suffix *-s*. *Bo·s* also means 'afterbirth' and 'navel' inter alia. Relevant Dongolese Nubian and Khalkha terms colexify 'house' with 'receptable' generally.

Other associations include: Efik *u'fök* is derived from *fok* 'to spread a covering over, to cover, disguise,' Hausa *'daki* is inter alia also used as a term for a year when counting the age of a horse, and Swahili *nyumba* contains *umba* 'to create.' Kwoma *aka* also means 'cave,' and *akama* also 'social role' and 'home region.' Kyaka *anda* also denotes an 'open valley area' as well as 'adobe' inter alia, Rotokas colexifies 'cabin,' Kosarek Yale *ae* also means 'region,' Yir Yoront *ngolt* also 'wall' of a house as seen from inside, Basque *etxe* also "lodging, shelter" and "frame, body" in the technical sense, and Nivkh *tyf* also means 'quarters.' Acoma *káçə*, when interpreted verbally, means 'it is tall,' Blackfoot colexifies 'house' with 'lodge,' and Ineseño Chumash *ma'm*, a rare word for 'house,' is also used adpositionally with the meaning 'inside of.' Lesser Antillean Creole French *kai* also means 'fishscales' inter alia, Tuscarora *unę́hseh* also is used with the meanings 'cage' and 'umbrella,' Yana *-sʒa-* also means 'upward,' and Hupda *mə̌y* also 'comb, brush.' Bororo colexifies 'house' with 'palm leaf' (similarly, Jarawara *yobe* is also the name of a palm species, and since Jarawara colexifies 'house' with 'roof, thatch' this may be the chain of associations that is also responsible for the association in Bororo), Embera with 'inn,' and Hupda with 'burrow' and 'brush.' The Maxakalí term *mĩp-tut* is analyzable as 'wood-mother' or 'wood-woven.net' (compare Cashinahua *jive* 'house, village, commuity' and *ji* 'tree, wood). Wayampi *-ɔka* as a verb means 'to cut with axe,' and *(l)ɛtã* also means 'site of waterloving creatures' ("Gîte des monstres aquaphiles"). Hani *laqhyul* contains *hyul*, meaning 'inside, domestic' (alongside 'to be extremely comfortable'), Lenakel *nimwa* also denotes the 'placenta, afterbirth,' a 'coccoon,' and a 'handle,' while Hawaiian *hale* also means 'institution' as well as 'host, hospitable person.' Finally, Bislama *haos* also denotes the 'bridge of a ship' or the 'cabin,' as of a truck, and Sedang colexifies 'house' and 'rainy season.'

82. The Key

Representation: 68%
Motivated: 37.0%
Thereof Analyzable: 33.0% Thereof Colexifying: 5.3%
Thereof by Contiguity: 33.5% Thereof by Similarity: 3.3%
Recurrent associated meanings: open/close, lock/keyhole, lock/unlock, door, house, tap

A common association for the 'key,' which is predominantly realized by morphologically complex expressions, is that with terms meaning either 'open' or 'close.' It comes in a variety of structural guises. Terms may be of the derived kind, as in Chukchi *ine-nwentet-*

icɣən 'ANTIPASS-open-INSTR,' occurring also in Khoekhoe, Hausa, Swahili (by prefixation of a noun class marker), Burarra, Welsh (where the relevant term also means 'opening' and 'act of opening'), Blackfoot, Chickasaw, Central Yup'ik (Nunivak island dialect), Cashinahua, Guaraní, Piro, and Fijian (straightforward evidence for this patterns from Indo-European is only found in Celtic, Buck 1949: 469). Terms of the lexical type include Yoruba *ì-ṣí-kà* 'NMLZ-open-thing' (this term is not usual), Biloxi *ti´ iⁿtpa´xoⁿni´* /ti iⁿ-dupaxi´-oⁿ-ni´/ 'house INSTR-open.door-make-CAUS,' and the analogous Kiliwa and Arabela terms *wa?=h+kap-u?* 'house=3+open-OBL' and *tiootiu riatataja* 'door opener.' In Comanche, there is an additional constituent making reference to a pointed object present, Maxakalí has *pipkup mõhãm-yĩnnĩn-'ax* 'metal.object ??-shut-can,' and Tehuelche *golk'o kotenwe*, *gonke:renwe*, *gonomk'enwe*, and *go:nko' kašomk'enwe*, all of which feature a term for 'door,' *gonk'o ~ go:nk'o*, and the instrument nominalizer *–we* alongside verbs meaning 'loosen,' 'shut,' 'open' and 'see' (a semianalyzable term with 'loosen' is also featured in Nez Perce). Hawaiian, due to the flexibility of its lexemes with respect to the syntactic slot they may occupy, colexifies 'key' with 'to open' inter alia. Further, Oneida has a term for 'key,' *atenhotukwátha?*, derived from a complex verb meaning 'to shut or close a door' specifically. Semianalyzable terms involving either 'open' or 'close, shut' occur in Khoekhoe and Cheyenne, and note also the similarity between Laz *nkula* 'key' and *nkol* 'close.' A related pattern is that when constituents mean 'lock' or 'unlock' rather than the more general 'open' and 'close;' in fact, Rotokas combines all those meanings in its term for 'key': *tupa karu-pa* 'close/lock open/unlock-DERIV.' Efik features a derived term (*u-kpähäre* 'NMLZ-unlock'), so do Khoekhoe (in which 'open' is also colexified with 'unlock'), Haida, and Piro; in Pawnee and Yuki, the term is of the lexical type (*rakcaakarikuku* /rak-caakarikuk-hus/ 'tree/wood-unlock-IPFV' and *ṗiṭ ṗiṭ-uƖ* 'door lock-INSTR'). Furthermore, Upper Chehalis has *s-q̓ólq̓=iyq* 'CONTINUATIVE-lock.up=house.' Semianalyzable terms involving a verb meaning 'to lock' are found in Kildin Saami, Hani (the other element means 'similar, true to life' when occurring on its own), and Yay. In a number of languages, 'key' is colexified with its counterpart, the 'lock' or 'keyhole.' This is the case in Hausa (where the relevant term is also dialectally used as the title of the principle advisor of the Emir inter alia), Nez Perce (by the term *weċéʔkeʔs,* containing *we-* 'hit' and *ċéʔ* "be trapped, be blocked in, be stranded, be struck, be immovable," colexifying 'hammer' additionally), Wappo, Wintu, dialectally in Central Yup'ik, and in Abipón. Similarly, Sko *long* also means 'hole, cave.'

As the discussion of terms with 'open' and 'close, shut' has shown, in many languages 'door' provides an additional contiguity anchor. Other terms with a constituent meaning 'door' are Abipón *l-aham-kate* 'POSS.INDEF/3SG-door-INSTR,' Bororo *baiporo epa* 'door instrument,' and Toba *lemaqte ñi lasom*, containing *emec* 'spoon' and *lasom* 'door.' In addition, complex terms involving a constituent meaning 'lock' are found in Miskito (*ki mita* 'lock hand'), Hani (*zovqdul*; *zovq* is 'lock' and *dul* can refer to a variety of longish objects, among them 'thigh' and 'pen'), Malagasy (*fanalahìdy*, analyzable as /fanàla-hidy/ 'instrument.to.take.out-lock'), as well as Yoruba (*ọmọ àgádágodo* 'child padlock') and Manange (*1tʌntsʌ-1amʌ* 'lock mother'). Extension of kinship semantics to the pair 'key' and 'lock' is also found in Latvian (Buck 1949: 469), compare also Matisoff (1992), who shows that it is widespread in Southeast Asia.

Furthermore, a Ngaanyatjarra avoidance register term alongside other meanings colexifies 'key' with 'knife,' and Basque and Santiago Mexquititlan Otomí colexify 'key' with 'tap' among other meanings (the latter language also has an optional complex term with *bojä* 'iron').

Other associations include: Muna *kunsi* also means 'button,' "have a commercial partnership," and 'plot, to conspire.' The Abzakh Adyghe term *ʔ°ənčʲəbze* is analyzable as /ʔ°ə<n>čʔə-bze/ 'abstract<RELAT>-strap.' Japanese *kagi* also means 'hook' (though the respective readings are distinguished in writing). Khalkha *tylki-gyr* is analyzable as 'push-INSTR' and also may refer to any "long-handled implement used to push snow, manure, etc." Nez Perce *wahołkaʔs* contains *we-* 'with an implement' and *hoł* 'to loosen,' and Nuuchahnulth *ƛ̓iqy̓ak* is analyzable as /ƛ̓iqʷ-y̓ak/ 'untie-tool.' Tuscarora *uhsęwáʔreh* also means 'fork,' 'nail' (on this origin for words for 'key' in Indo-European see Buck 1949: 468) and 'needle,' and Copainalá Zoque *wi'toquiuy* contains *wi'tu* 'to turn, be locked with key.' Huambisa *yawi* is also used to refer to 'safety pins,' wheras Miskito *warbaika*, analyzable as /warb-aia-ka/ 'turn-INF-DERIV' also means 'tongs.' Fijian *kī* also denotes a fish species and means 'to do' (the meaning 'key' is due to borrowing from English), Malagasy colexifies 'needle of pine tree,' Rotuman *kī* also means 'sepia' inter alia, while Sedang *khúang* also means 'drill' and 'brace.' Bislama *ki* (< Engl. *key*) also means 'gear' and 'spanner,' as well as 'key' in the music-related sense and denotes the 'pegheads' of a guitar.

83. The Knife

Representation: 94%
Motivated: 20.9%
Thereof Analyzable: 18.7% Thereof Colexifying: 2.5%
Thereof by Contiguity: 14.6% Thereof by Similarity: 4.1%
Recurrent associated meanings: machete, cutting implement, cut, iron/metal/steel, sword, bamboo, razor, to skin, fish

As with other artifacts, terms derived from a verb denoting the principal activity that can be performed with that artifact, in this case 'to cut,' are also frequent for the meaning 'knife' (see also Buck 1949: 558 for Indo-European), but clearly not as frequent as with other artifacts, presumably due to the fact that knifes are frequently indigenous tools and were present before the era of colonization. Derived terms are nevertheless found in Chickasaw, Kashaya, Abipón (where 'to cut' is colexified with 'to shine'), Bora, and Ancash Quechua. Fijian has a derived term from a verb meaning 'to cut with knife' specifically, Khoekhoe has an obsolete term of this kind, and the association is recoverable etymologically for Dongolese Nubian, Kolyma Yukaghir, and Kiowa. Furthermore, San Mateo del Mar *nicojchay onij* contains *acooch* 'cut' and *onij* 'meat' (the language also has another term, *nitajcüy cüet* containing *ataag* 'disembowel' and *cüet* 'fish,' for which compare Yir Yoront *ngartyann* containing *ngart* 'fish' and *ye* 'cut, slice'). In contrast, Toba *lpetegaxanaxat* contains the verb *petec*, meaning 'to cut hair' specifically (this term colexifies both 'scissors' and 'razor,' which latter association is shared by Nez Perce and Tuscarora). The meanings 'knife' and 'cut' are colexified in Bwe Karen and Samoan (here, by a polite term), and syn-

chronically semianalyzable terms of this kind are found in Kolyma Yukaghir, Kaingang, and Wayampi. Badaga, Sora, Cahuilla, Lake Miwok, Nez Perce, and Mandarin colexify 'knife' with 'sword.' Colexification with other cutting implements (such as a 'sickle,' a 'dagger,' or an 'axe'), or general terms for any sort of bladed tool is also found in other languages, namely Badaga, Khalkha, Lake Miwok, Nez Perce, Oneida, Rama, Toba, and Samoan. More specifically, Meyah, Sko, Basque, Cubeo, Hupda, and Jarawara colexify 'machete,' and four languages of South America also have complex terms on the basis of 'machete': Bora *nɨ́ɨtsúwá-wu* 'machete-DIM' (in fact, the word for 'machete' is derived form a verb meaning 'to cut'), Cavineña *cuchiro caca* 'machete small,' Chayahuita *cosëra'hua* ~ *cosora'hua*, probably containing *cosoro* 'machete for cutting grass' and a classifier suffix, and Yanomámi *sipara si* 'machete cover.' Six sampled languages, Baruya, Cahuilla, Kiliwa, Central Yup'ik, Kaingang, and Lengua colexify 'knife' with the material 'iron' or 'metal' and/or 'steel' (Baruya also with 'plastic),' while Cubeo has *tãu-ve* 'glass/metal-CLASS.SLIM.SLENDER.FLAT.OBJECT' and Toba *laicaua laỹi* 'metal/iron edge.' Four languages of broader Oceania, Kwoma, Lavukaleve, Lenakel, and Samoan (by a polite term) colexify, by provenience contiguity, 'knife' with 'bamboo' (Lenakel also with 'backbone' inter alia, and Samoan also with 'fishing rod'). In Kiliwa, the word for 'knife' is derived from a verb meaning 'to skin' (*na(y)=c+ruuw* 'child/small=INST/mouth+to.skin'), and Hawaiian colexifies these meanings alongside 'flint' and "to stand on edge."

Other associations include: Khoekhoe *ǂnamib* ~ *ǂnamis* 'simple hand-made knife' is related to *ǂnami*, a root occurring in the Damara dialect and meaning 'to chip, chop off' inter alia. Koyraboro Senni *huri* also dialectally means 'seek, look for,' while *zaama* dialectally is also a particle meaning 'because.' Ngambay *kìya* also means 'to hide,' 'to place, set.' The Burarra term *angujarrcha* contains *jarrcha* 'to slice off,' while Dadibi *ge hwa* contains *ge*, meaning 'nut, egg' or 'small object' in general. Ngaanyatjarra *kunmarnu* is an avoidance register term for 'Sunday, week, jumper,' 'key,' and 'knife.' Rotokas *visi-paa* is analyzable as /visi-pa/ 'poke/hit-DERIV.' Toaripi colexifies "tying, lashing material" inter alia, and Badaga *su:ri* also means 'sharp, pointed' in an adjectival sense. Japanese *hō-chō* is analyzable as 'kitchen-man.' Haida *q'it'uhl 'laʔáaw* contains the instrumental prefix *q'it'-* 'cut with knife' and presumably *uhl* 'shape.' Itzaj *lomik* contains *lom* 'to stab,' and also denotes a 'stab,' 'stake,' or 'nail.' The Kashaya term *qahca* also means 'missile' and 'clitoris.' Nuuchahnulth *ʕak-ẏak* is analyzable as /ʕak^{w}-ẏak/ 'whittle-tool,' and Wintu *ċebet* 'stone knife' is related to *ċEb-* 'sharp, knifelike, plane, whittle.' The Arabela term *cushiishi* contains *cushi* 'pig' (perhaps folk etymology of Span. *cuchillo*?). Maxakalí *mĩkax* also means 'rock, stone,' Bislama *naef* also 'blade' specifically, and Rotuman colexifies 'knife' with 'to circumcise' inter alia. Samoan has *fa'aola fanua* 'save/savior land/field' for 'adze, axe, knife' (the term is restricted to polite usage), and Sedang *rơkong* is also used with the meaning 'mouth, language, word.'

84. The Ladder

Representation: 78%
Motivated: 52.9%
Thereof Analyzable: 33.3% Thereof Colexifying: 20.0%
Thereof by Contiguity: 24.2% Thereof by Similarity: 25.4%
Recurrent associated meanings: stairs/staircase, climb/ascend, step, bridge, wood/tree, foot, scale, road/path, leg, hole, thing, sky, walk, lean

Frequently, terms for this meaning are of the derived kind, the derivation base being verbs meaning 'to climb,' 'go up,' or 'ascend,' as in Rotokas *iipa-pa* 'climb/go.upward-DERIV.' Alongside semianalyzable terms in Ineseño Chumash, Hani, and Kapingamarangi, this is realized by derived terms also in Yoruba, Muna, Sora, Blackfoot, Upper Chehalis, Carrier, Cheyenne, Chickasaw, Kiliwa, Nez Perce, Wintu, Central Yup'ik (colexifying "neck opening of parka"), Aguaruna, Bora, Chayahuita, Guaraní, Fijian, and Rotuman. There are also terms with an additional element bearing lexical meaning. Mbum has *fè-hénà* 'thing-climb' (a similar term with 'thing' is also featured in Hupda), Khoekhoe has *!apa-haib* 'climb/ascend-stick/tree' for a traditional type of ladder. Somewhat differently, Carrier has *uḳwe-tera-î-thi* 'on.it-getting-up-road' (complex terms with 'path' are also Lengua *amai letin* /amai netin/ 'path sky/above,' which colexifies 'ladder' with 'dam' as well as Hawaiian *ala-pi'i* 'path-climb,' which denotes 'ladder' as well as 'step,' while the association is realized by colexification, also with 'door' and 'gate,' in Haida). Moreover, Piro has *hatsko-pi-xe* 'ascend-rod-pole,' and similarly, Blackfoot has *iihtáísokamisáóo'p* /iiht-á-sok-wamis-oo-o'p/ 'INSTR-DUR-above-??-go-21.NOM.' This term, like the Hawaiian one, colexifies 'ladder' with 'step,' and this is indeed a very frequent association in the languages of the sample, and some of the terms derived from 'to climb' above exhibit this pattern as well. It is also found in Dongolese Nubian, Kaluli, Lavukaleve, One (here also with the meaning "horizontal brace of a pangal bed"), Sahu, Toaripi, Khalkha (also with 'footboard' and 'pedal'), Upper Chehalis, Arabela, Cayapa, Bislama, Fijian, and Sedang, which also colexifies 'stem' (the original meaning of Japanese *hashigo* is also 'step' diachronically). Similarly, Kaingang has a term for 'ladder' derived from a verb meaning 'to step.' A semianalyzable term for 'ladder' where one of the constituents means 'step' is found in Wayampi, and one where it means 'to step on, set foot on' in Kwoma. Another common pattern of colexification is that with 'stairs' and/or 'staircase,' found in Efik, Dongolese Nubian, Yoruba, Berik, Muna, Rotokas, Sentani, Toaripi, Basque, Khalkha, Blackfoot, Upper Chehalis, Chickasaw (which also colexifies 'fire escape' specifically), Comanche, Nez Perce, Pawnee, Central Yup'ik (colexifying also 'rung'), Embera, Jarawara, Maxakalí, Hani, Hawaiian, Kapingamarangi, Manange, Rotuman, Bislama, Fijian, Manange, and Mandarin.

Furthermore, Basque has *esku-eskailera* 'hand-stairs,' and Khalkha *giški-gyr* 'step-bridge;' 'bridge' (or specific types of bridges) and 'ladder' are furthermore colexified in Carib (which also colexifies 'harbor'), Lengua, Toba, Yanomámi (where the relevant term *ihirakɨ* is derived form *ihira* 'to construct a frame' by means of suffixation of the quantal classifier *-kɨ*, for which see § 4.4.1.1.), Hawaiian ('plank bridge' more specifically, and also colexifying 'trestle'), and Sedang. Five sampled languages have complex terms in which

one of the constitutents means 'foot': Efik *udik'uküt* /u-dik'hi-uküt'/ 'NMLZ-tread.upon-foot/leg,' Abzakh Adyghe *λeywen* /λ(e)-ye-we-n/ 'foot-UPWARD.MOTION-stomp-HAVE.INTENTION.TO,' Kiowa *'ą̀n-t'out-'ʜ'dɑ* containing *'ą̀n* 'foot' and *'ʜ'dɑ* 'pole' (compare also *'ą̀n-t'out* 'to climb up steps'), Miskito *minamangka*, containing *mina* 'foot' and *mang* 'to put' (there is a further semianalyzable term with *mang* in this language), and Bwe Karen *kha-bɔ́* 'foot/leg-handle/holder.' Further complex terms involving 'wood' or 'tree' are Biloxi *a´ya^{n} i^{n}de´ ~ aya´i^{n}de* /a´ya^{n}´ i^{n}-de/ 'wood INSTR-go,' Tuscarora *urę?náhrareh*, consisting alongside grammatical material of the verb-*ahrą̄r-* 'be a hole' and the incorporate *-rę́T-* 'tree, log' (compare colexification of 'hole' and 'ladder' inter alia in Hawaiian), Maxakalí *mĩp-ku'ĩn* 'wood-slashes/stripes,' and Ket *bulaŋdoks* /būl-aŋ-d-ōks/ 'leg-PL-POSS-wood' (for the association with 'leg' compare also Efik *udüri-uküt* 'top-leg'). Note also that the Cashahuita term is derived from a verb meaning 'to climb' by a classifier for wood-related items, and that there are semianalyzable term with 'wood' in Basque and Aguaruna.

Efik *ë-beri* is derived from *beri* 'to lean,' and a term where the meaning 'leaning' figures is also found in Pawnee (an optional complex term with a verb meaning 'to lean up' is also found in Haida on the basis of the term colexifying 'door,' 'gate,' and 'ladder' mentioned above). Alongside Lengua *amai letin* 'path sky/above,' which was already mentioned above, Kiliwa also has a term betraying an association with sky: *?-mai?=t-h-?+paa-y-u?* is analyzable as 'DN-sky/heaven=SUBJ-3-DN+depart-ATT-OBL.' Dadibi *togobili* may be analyzable as /togobe-bilibo/ 'edgeposts-walk,' and Abipón *n-acaR-haR-late* as 'POSS.INDEF/3SG-walk-the.one.who-LOC.' This term also means 'shoe' (compare section 91). Finally, Itzaj, Embera, Fijian, and Hawaiian colexify 'ladder' with 'scale,' and Rotuman colexifies 'ladder' also with 'stave, staff' and also music written in this notation, and Hawaiian with 'scale' in music.

Other associations include: Buli *tiili* colexifies 'ladder' with "the clay connection between two ancestral shrines." Hausa *tsani* also denotes an 'intermediary' inter alia, and Kanuri *kùrángá* also means 'monkey.' Ngambay *ddíki* is also used to refer to "something difficult," and *mbata* colexifies 'stool.' Yoruba *àkàbà* contains *bà* 'perch on, alight' and the nominalizer *á-*. Baruya *kwaaka* also means 'soil, earth, dirt, clay, ground, land, country' as well as "shooting short of a target," while Gurindji *tankuj* also may refer to a 'useful thing.' Kwoma *akatoko* (containing *aka* 'house') is also used with the meaning 'escalator,' and *piitiishey* also means 'scaffolding.' Muna *lawa* also means to "answer back, respond" as a verb, and for *pulangku ~ polangku*, compare *langku*, which denotes a part of a loom, but also means "social rank, level." Ngaanyatjarra *lata* also means 'letter' (due to collapse of Engl. *ladder* and *letter*). Kosarek Yale *modobak* is derived from the verb *modob-* 'walk or climb using footholds' and colexifies 'foothold' as well as 'pole with notches.' Badaga *so:pa:na* is also the name of a "log in which steps have been cut," Khalkha *šatu(n)* also means "phase, stratum, level," and Welsh *ysgol* also means 'school.' Oneida *yelathʌstákhwaʔ* is analyzable as /ye-lathʌst-hkw-haʔ/ 'FEM.INDEF.SG.AGENT-get.something.up-INSTR-HAB,' Central Yup'ik *akeq* colexifies 'ladder' with 'barb' and 'rung,' and Bora features also a term on the basis of a verb meaning 'to descend' rather than 'to ascend' as reported above: *niityé-wááhyo* is analyzable as 'descend-CL.layered.things.' Cashinahua *tapaiti* is derived from *tapa* 'floor' by means of the instrumental suffix *-ti*. Jarawara *ki-kisima* is analyzable as 'RED-come.down.'

Toba *piaxalate ~ napiaxalate* also means 'pulpit,' and the Tsafiki term *teranca* appears to contain *terano* 'to dance' and *ca* 'in front of.' Bislama *step* also conveys the meanings 'terrace on cliff' and 'gait, pace,' while Hawaiian *haka* also means 'platform, shelf' inter alia, and Manange *2li* also 'face.'

85. The Mirror

Representation: 77%
Motivated: 51.3%
Thereof Analyzable: 34.5% Thereof Colexifying: 16.8%
Thereof by Contiguity: 24.5% Thereof by Similarity: 5.3%
Recurrent associated meanings: glass/type of glass, see, look/inspect, shadow, window, reflect/reflection, water, face, eyeglasses, eye, bright/brighten, thing

As for virtually all artifacts, terms making reference to the purpose they serve abound. In this case, this means that very frequently, terms make reference to seeing, looking (see Buck 1949: 454 for Indo-European), or reflecting. Derived terms on the basis of verbs meaning 'to see' frequently also contain a reflexive marker (as did a Sanskrit term, Buck 1949: 454), as in Chickasaw *aailipisa*, analyzable as /aa-ili-pisa-'/ 'LOC-REFLX-see-NMLZ.' There are also derived terms in Mali (where the meaning of the derivation base is 'see past' more precisely), Upper Chehalis, Comanche, Kashaya, and Nez Perce. In lexical terms, 'shadow' is often the meaning of the additional lexical element (an association likewise reported for Indo-European by Buck 1949: 454), as in Mbum *ákó-têm* 'see-shadow,' and also in Japanese and Pawnee (associations with 'shadow' also occur in other configurations, by direct colexification in Lavukaleve, here also with 'spirit,' Rama, also with 'picture,' as well as Sedang, and by the analyzable Chukchi term *wiilyiteneŋ* containing *wiil* 'shadow' and *ɣite* 'watch' and Samoan *fa'a-ata* 'CAUS-shadow/image'). Otherwise, Efik has *u-kur-isü* 'NMLZ-see-face,' and Yir Yoront *kowllewkarrlnh* and *kowllewkerr(w)lh* contain *kowllew* 'face' and *karr* 'see, look at' and furthermore colexify 'mirror' with 'photograph.' Biloxi *o^{n}´do^{n}ho^{n}ni´* revolves around the root *do^{n}* 'to look at, see,' Cheyenne *amôhóomâhtsestôtse* contains *amôhóom* 'to see in reflection' and *hestôtse* 'thing' (and indeed also means 'reflection'), and semianalyzable terms with 'see' occur in Buli, Hausa, Tehuelche, and Great Andamanese. Terms derived from verbs meaning 'to look' or 'inspect' are found in Kanuri, Rotokas, Kiowa, Oneida and Fijian, where the derivation base means 'to look at, as a reflection in water or in a mirror' specifically. Rotuman colexifies 'mirror' with "to watch closely, gaze at" directly. Furthermore, Yoruba has *à-wò-jìjìí* 'NMLZ-look-shadow/reflection,' Nivkh *un'ɣr-njus* 'star-place.to.look,' and Carrier *pê-na-tsẹ-n-de-nel·ên* is analyzable as 'wherewith-repeatedly-??-rotundity-at.one's.own-look,' with the element glossed as 'rotundity' referring to the 'face' (other complex terms more or less tightly connected with 'face' other than those already mentioned are Itzaj *eetz'-'ich* 'face-make' ('face' in the sense of 'grimace,' though), Cashinahua *beisikiti*, presumably analyzable as /beisikiki-ti/ 'look.at.other.person's.face-INSTR' and Cubeo *jiva-rɯ* 'face-CLASS.ROUNDISH.OBJECT). Terms derived from verbs meaning 'to reflect' or nouns meaning 'reflection' are found in the Nunivak Island dialect of Central Yup'ik, where the relevant term *tarenriurun* is analyzable

as /tarenriur-(u)n/ 'look.at.one's.reflection-device.for,' Malagasy (*fi-tàratra* 'NMLZ-light.beam/reflection'), and also in Haida and Tetun, while Kashaya *ʔama puṭʰam* contains *ʔama* 'thing' and *ṭʰaᵒm* 'to reflect;' Wintu *ʔilʔiloqma* can also refer to "anything that reflects; something shiny."

A pattern that is also widespread in the languages of the world is colexification of 'mirror' with 'glass' or types of glass, such as colexification of 'mirror' with 'pane of glass' specifically in Swahili. Such associations occur also in Hausa, Yoruba, Burarra, Kyaka, Muna, Ngaanyatjarra, Abzakh Adyghe, Bezhta, Upper Chehalis, Highland Chontal, Itzaj, Pawnee, Tuscarora, Great Andamanese, Fijian, Hawaiian (also colexifying 'clear, transparent, obvious' and 'cool' among other meanings), Bwe Karen, Malagasy, Sedang, White Hmong, Rotuman, Takia, Yay, and Bislama (note that according to Buck 1949: 454, this pattern within Indo-European "seems to be peculiar to English"). White Hmong, in addition, has *tsom-iav* 'look-glass.' Due to this general pattern of colexification, Hausa, Bislama, and White Hmong also colexify 'mirror' with 'eyeglasses.' Furthermore, Upper Chehalis, Kashaya, Kiowa, Pawnee, Tuscarora, Wintu (by the term *kenwiˑnaˑs*, containing *ken* meaning inter alia 'in, be in, put in' and presumably *wiˑ* 'male;' there is a reconstructed root **wi* meaning 'person' more generally), and Fijian colexify 'mirror' with 'window,' and there are other similar scattered patterns of colexification, such as for instance that with 'telescope,' 'binoculars,' 'thermometer,' and 'barometer' in Bislama, which also colexifies 'car window' specifically.

Another rather unrelated pattern is the association between 'mirror' and 'water,' due to the reflecting properties of water. This association, in the form of a pathway of semantic extension under circumstances of acculturation, is hinted to by the fact that Tuscarora *awéˑkyeh*, meaning 'glass,' 'window,' and 'mirror,' also means 'liquid.' Fijian also directly colexifies 'mirror' with 'water' by the analyzable term *i iloilo* 'DERIV look.at.reflection-RED,' Santiago Mexquitilan Otomí colexifies 'mirror' with 'river,' Noni has *me ɛ joo* '1SG in water' (the term also denotes a 'riddle'), Anggor *hoe-hɨmbo* 'water-ear,' Toaripi *ma ove* 'water picture' (lexicographer's comment: "before glass or metal mirrors were available, reflections in water served as a mirror"), and Kiliwa *xaʔ=ny-ʔ+yuw-l=t-p+sa-w-uʔ* 'water=POSS-DN+eye-ILL=OBJ-MP+see-DIR-OBJ.' In Buin, *tiinura* is a term for "water collected in trees" that is used as a mirror, and Berik has a semianalyzable term. The association with 'eye' as found in Kiliwa is also present in Ket (*deskəŋejroks* /dēs-kəŋej-r-oks/ 'eye-??-POSS-wood'), and semianalyzable terms where one of the constituents is 'eye' are featured in Waris (where this term is archaic) and Badaga. Finally, Kashaya has *ʔama piṭʰam*, containing *ʔama* 'thing' and *-ṭʰaᵒl-* 'bright, light enough to see by,' and similarly, Yanomámi has *warara-rimɨ* 'brighten.up-NMLZ'

Other associations include: Hausa *madubi* and *soka* also mean 'sorcerer' and 'wizard,' Muna *paeasa* assumes the meaning to "follow the example" when used verbally, and Basque *ispilu* also means 'white spot, white patch.' Khalkha *toli(n)* is also used with the meanings 'metal plate' and 'dictionary, vocabulary,' and Welsh *drych* also means 'sight.' Blackfoot *sáapia'tsis* contains *sap* 'in, within' and the instrument nominalizer *-a'tsis*, Kiliwa *x+ʔal=h-haʔ-uʔ* is analyzable as 'CAUS+??=3-seek-OBL,' Nuuchahnulth *piiḥẏaksaɬa* as /piḥ-ẏaksaɬa/ 'observe/study/examine-tool-at.or.on.the.forehead,' while Santiago Mexquititlan

Otomí *hñe* also means 'to put on' and San Lucas Quiaviní Zapotec *gyigwàan* also 'very beautiful girl.' Piro *haniha-pi* is analyzable as 'shade.eyes-rod,' and Wichí *to-peyak-hi* as 'POSS.INDEF-image-LOC.in,' while Hawaiian *aniani* appears to be reduplicated from *ani*, one of the meanings of which is 'for a hand to pass over a surface.'

86. The Needle

Representation: 90%
Motivated: 35.8%
Thereof Analyzable: 14.5% Thereof Colexifying: 21.6%
Thereof by Contiguity: 12.2% Thereof by Similarity: 11.0%
Recurrent associated meanings: injection/syringe, sew, pin, thorn, nail, hand of clock, thread/string, bone, cloth, fork, net shuttle, pierce, thing

For 'needle' (or 'awl'), the most frequent association is unsurprisingly that with 'to sew' (see also Buck 1949: 412), by terms of the derived type for in Central Yup'ik (which has, for instance, *mingqun* /mingqe-(u)n/ 'sew-device.for'), Upper Chehalis, Chickasaw, Kiliwa, Arabela, Fijian (where *cula* not only means 'to sew' but also 'to pierce, let blood, vaccinate,' for this compare Nuuchahnulth *qačak* /qač-akʷ/ 'pierce/puncture-tool'), and Malagasy. Terms of the lexical kind have 'thing' as the meaning of the additional constituent in Koyraboro Senni *taa-haa* 'sew-thing' (which is a substitute term for a monomorphemic term for 'needle' that must not be uttered at night), 'cloth' in Comanche *wana tsahkʉna?* /wana tsahkʉnarʉ-?/ 'cloth sew-NMLZ,' and 'bone' in Kapingamarangi *iwi due mee* 'bone sew thing' (semianalyzable terms containing verbs meaning 'to sew' are attested in Nuuchahnulth, Miskito, and Tehuelche, and there is one containing 'thing' in Dadibi). Both associations just mentioned are also found in different configurations. Buli has *garupein* /garuk-pein/ 'cloth-arrow,' Yir Yoront *mipkallnh*, containing *mip* 'cloth' and *kal*, meaning inter alia 'spear, poke, pierce, cut,' Rotuman *sui susuag hạ'u* /sui susuga hạ'u/ 'bone sewing clothes,' and a semianalyzable term where the identifiable constituent can be identified to mean 'veil' is found in Greek. Wappo and Yuki exhibit the association with 'bone' by colexification. In Upper Chehalis and Rama, 'needle' and 'thread' are colexified (in Rama 'silk grass needle' more specifically), while Lengua has *sokyi-tama* 'carry-string' and Piro *tsa-pu* 'thread-ENTITY.' Similarly, Hawaiian colexifies 'needle' with verbal 'to thread beads' and 'to string pierced objects.' Presumably by provenience contiguity, Anggor, Baruya, Buin, Khalkha, Abipón, Bora, Chayahuita, Tehuelche, and Yanomámi colexify 'needle' with 'thorn' (Yanomámi also with 'porcupine quill'), and Hupda has *măc-ʔut* 'metal-thorn.' Samoan colexifies 'needle' with 'sting' (see Buck 1949: 412 for cognates evidencing this association by semantic shift in Indo-Euopean). There are also patterns of colexification due to semantic extension. Basque, Khalkha, Lesser Antillean Creole French, and Tetun colexify 'needle' with 'hand of clock,' and Koyraboro Senni, Ngambay, Noni, Rendille, Swahili, Anggor, Baruya, Bezhta, Khalkha, and Cubeo with 'injection, syringe.' Kwoma, Rotokas, Oneida, Tuscarora, Miskito, Tehuelche, and Hawaiian colexify 'needle' with 'nail' (Oneida also with 'wire,' and Tuscarora, Miskito, and Hawaiian also with many other objects made of metal, in Tuscarora for example 'auger,' 'fork' – this pattern of colexification is shared

with Nez Perce and Nuuchahnulth – 'key,' 'pitchfork' and "dwarf pike"), and Xicotepec de Juárez Totonac *līxtocon* is analyzable as /lī-xtokó-n/ 'INSTR-to.nail-INSTR.' Koyraboro Senni, Rotokas, Basque, Ket, Kiowa, Oneida, Cayapa, Hawaiian, and White Hmong colexify 'pin,' and San Mateo del Mar Huave and Nez Perce colexify 'needle' with 'net shuttle.'

Other associations include: Ngambay *énemé* also denotes a "kind of drill to make a hole with," and *suwa* also means "boil leaves rapidly in water, cook meat." Kyaka *wamyali ~ wamalyi* also means 'skewer.' Basque *orratz* is also used with the meanings 'needle of grammophone,' 'compass,' and 'switch,' while the denotational range of the Ket term *ìn* also includes 'fingernail' and 'claw.' Khalkha *zegyn ~ zegyy* also means 'East, oriental, left,' and Japanese *hari* also means 'beam,' with the terms being prosodically different. Biloxi *a^{n}sadûki* contains the word for 'pine tree,' whereas Cheyenne *vé'ho'êškóovo'hestôtse* is analyzable as /vé'ho'é-škóovo'hestôtse/ 'whiteman-toothpick.' San Mateo del Mar Huave colexifies 'needle' with 'snake, worm,' Itzaj *puutz'* also means 'enemy,' and Kiliwa *t-x-pa?-p-u?* is analyzable as 'OBJ-CAUS-possess.round.object-MP-OBL.' Wintu *ċu·p* also denotes the concept 'dagger' and anything with a sharp point in general. The Yaqui term *ji'ikiam* is analyzable as /ji'ik-ia-im/ 'weave-NMLZ-PL.' Abipón *n-icir-en-kat*e (variant form *nicirenRat*) is analyzable as 'POSS.INDEF/3SG-unite-VOL-INSTR,' and Bororo *akigu iĉira* as 'thread palm.fibre.' Hawaiian *pahele*, also denoting a 'snare, noose' or 'trap' as well as 'deceit, treachery,' seems to be derived from *hele*, meaning 'to tie, bind, lash, snare, noose' inter alia. Another Hawaiian varies somewhat in form depending on the variety of Hawaiian; that spoken on the island of Hawai'i, *hānai*, also means 'foster child, stepchild' inter alia. Finally, Manange *4t^{h}e* also means 'to hear,' Mandarin *zhen1*, going back to distinct Early Middle Chinese terms (Pulleyblank 1991: 401), also "true, exactly," and Yay colexifies 'needle' with 'gold.'

87. The Paper

Representation: 84%
Motivated: 59.6%
Thereof Analyzable: 13.2% Thereof Colexifying: 46.8%
Thereof by Contiguity: 48.0% Thereof by Similarity: 5.2%
Recurrent associated meanings: letter/book/document, write/draw, leaf, talk/speak, playing card, cloth, map, mail, role, cardboard, banknote, skin, wrap/wrapping

The most frequent association is, by contiguity, colexification with 'letter,' 'book,' or another type of document and sometimes also 'page.' It is found in Buli, Efik, Hausa, Koyraboro Senni (colexifying also 'amulet, talisman'), Mbum, Noni, Dongolese Nubian, Rendille, Yoruba, Buin, Burarra, Kwoma, Yir Yoront, Abzakh Adyghe, Badaga, Basque, Ket, Biloxi, Cahuilla, Upper Chehalis, Cheyenne, Chickasaw, Highland Chontal, Comanche, Haida, San Mateo del Mar Huave, Itzaj, Kashaya, Lake Miwok, Lesser Antillean Creole French, Nez Perce, Oneida, Santiago Mexquititlan Otomí, Pawnee, Pipil, Tuscarora, Central Yup'ik, San Lucas Quiaviní Zapotec, Abipón, Aguaruna, Bora, Bororo, Cashinahua, Cavineña, Cayapa, Chayahuita, Guaraní, Huambisa, Hupda, Kaingang (where *vẽnh ra* seems to contain *vẽnh* 'small plants'), Macaguán, Maxakalí, Miskito, Piro, Rama, Sáliba, Wayampi,

Yanomámi, Bislama, Hani, Malagasy, White Hmong, and Samoan. The relationship between 'letter' and 'paper' is marked by gender alternation in Embera, and that between 'sheet, piece of paper' and 'book, letter' in Khoekhoe by alternation of nominal designants. More specifically, Greek, Hawaiian, and Samoan colexify 'paper' with 'playing card' (Samoan also with 'banknote' and 'ticket'), Greek and Haida with 'map,' Nez Perce and Central Yup'ik with 'mail' (the Nez Perce term also denotes the 'Bible' specifically), Basque and Itzaj with 'role,' Abzakh Adyghe and Samoan with 'banknote,' and Cashinahua and Rotuman with 'cardboard.'

Otherwise, complex terms of the derived type from a verb meaning 'to write' or 'to draw' occur in Efik (*ñ-wet* 'NMLZ-write/mark/paint;' this term also denotes a 'printing, inscription' and 'form,' 'impression,' 'representation,' 'reflection,' and 'shadow'), Sora (*id'o:l-ən* 'write-N.SFX'), Arabela (*naajio-jua* 'writing-CLASS.GROUND;' this term also denotes a 'scribe' and a table used to write on), Abipón (*elerk-a* 'writing/letter-PL'), and Yanomámi. Indeed, the Abzakh Adyghe and Oneida terms colexifying 'paper' with 'letter' and 'book' are also of this type (the Chukchi, Cheyenne and Yanomámi colexifying terms are semianalyzable). The Yanomámi term colexifies 'paper' with 'pen.' Moreover, Tsafiki has *pilá quidó* 'writing/book skin/bark' (note also the colexification of 'skin,' 'hide,' and 'paper' in Buli and the origins of Indo-European words for 'paper,' Buck 1949: 1289), and Chickasaw *holisso* also means 'to be written' in verbal usage. Kwoma, Rotokas, Bezhta, Hupda, and Piro colexify 'paper' with 'leaf' (Kwoma with "dry banana leaves" specifically), as did Sanskrit (Buck 1949: 1289); note also Kaluli *mo:fo:s*, containing *fo:s* 'leaf' and perhaps *mo:* "base of tree stump or trunk," "basis or reason for utterance." In two languages of Eastern North America, there are complex terms for 'paper' where one of the constituents means 'cloth:' Lakhota *mniȟúha-khakháka* 'cloth-rustle' and Pawnee *raawihaakaraaʔiit*, which is analyzable as /raawir-taakaar-raaʔiit/ 'cloth-white-telling.' Moreover, Nuuchahnulth colexifes 'cloth' with 'paper;' the relevant term is *qicaaɬ,* analyzable as /qic-aˇɬ/ 'mark/paint/tattoo-on.a.fabriclike.surface.' The Pawnee term, betraying an association with uttering words, has a parallel in Nez Perce (*tíˑmi-s* 'speak-AGT') and in Abipón, where there is a word for 'paper' (as well as 'word') derived from a verb meaning 'to talk' (see also Buck 1949: 1003 for this association in Ancient Greek). Baruya colexifies 'paper' with 'wrapping,' and, analogously, Kiowa *mạtsêḥẹ-mạ* 'sheet of paper' is analyzable as 'wrap-NOUN.POSTFIX.' Bororo *bapera* also denotes leaves from the stomach of ruminants; Kiliwa *txpha?* may also perhaps refer to a "part of the ruminant digestive system called 'the book' (Sp. el libro), or related to sheepskin parchment," and similarly, Lake Miwok *pápel* also denotes "an internal organ attached to the stomach of animals which opens like a book when one cleans it." Since the Lake Miwok term is a borrowing from Spanish and the Kiliwa gloss refers to Spanish, it seems likely that this is a pattern copied from (local) Spanish.

Other associations include: Bakueri *liw̲aw̲é̲* also means 'wing,' Buli *gbang* also 'gambling,' and Rendille *khadáab* also 'scab.' Burarra *(-)jurra* colexifies 'paper' with 'track, footprints,' and Kyaka *pepa* may also refer to 'stationry' or 'tissue.' Yir Yoront *waqrr* also denotes the bark of the tea tree or melaleuca, and Japanese *kami* also denotes the 'hair' and 'God,' with the individual meanings distinguished in writing. Biloxi *akŭtxyí* contains the word for 'spotted, striped,' and the Carrier term *etesţ̀es* also means 'fur.' Pipil *a:mat* is

also the name of the amate, a species of fig tree, Central Yup'ik *igaq* also means 'mark, symbol' and 'letter of alphabet,' and Bora *waajácuháámi* is analyzable as /waajácu-ʔa:mɨ/ 'study/know/understand-SCM.leaf.' Rotuman *pepa* also means 'pepper' (< Engl. *paper* and *pepper* respectively).

88. The Pen

Representation: 66%
Motivated: 51.6%
Therof Analyzable: 45.4% Thereof Colexifying: 6.2%
Thereof by Contiguity: 39.7% Thereof by Similarity: 3.4%
Recurrent associated meanings: write, stick, feather, paper, wood/tree, crayon, ink, rod, chalk, paint, brush, typewriter, thing, poke, bone

Terms for 'pen' frequently contain verbs meaning 'to write,' which may in turn themselves have experienced semantic extension from 'to scratch,' 'to mark,' or like meanings which are not discussed separately here (the association between 'pen' and 'writing' is quite rare in the diachrony of Indo-European, mentioned only for Sanskrit by Buck 1949: 1290). Derived terms, such as Blackfoot *iihtáísínaakio'p* /iiht-á-sínaaki:??-o'p/ 'INSTR-DUR-write:??-21.NOM' are featured in Rotokas, Upper Chehalis, Chickasaw, Kashaya, Lake Miwok, Nez Perce (colexifying 'seal'), Oneida, Quileute, Xicotepec de Juárez Totonac, Tuscarora, Yuki, Central Yup'ik, Aguaruna, Arabela, Bora, Carib, Cashinahua, Cubeo, Guaraní, Kaingang, Piro, Imbabura Quechua, Toba, and Yanomámi. Variants of this pattern are found in Haida, where the derivation base of a relevant term (colexifying 'chalk' and 'crayon,' a pattern of colexification shared with Piro, and with Lesser Antillean Creole French in the case of 'crayon') means 'to write a letter' specifically. Cheyenne and Mbum have analyzable terms of the lexical type where the second constituent means 'thing.' Otherwise, 'wood' or 'tree' is a frequent meaning for the second constituent, as in Pawnee *raktariihkaruukus*, analyzable as /rak-rariihkaraʔuk-hus/ 'tree/wood-write-IPFV.' Such terms are also found in Ngambay (where 'tree, wood' is colexified with 'stick' inter alia), Kiowa (colexifying 'writing table'), Maxakalí, Tsafiki (in the latter two languages, 'tree' is colexified with 'bone'), and Yir Yoront, where the relevant term *yoqlatkallnh* contains *yoq* 'tree, stick, wood,' *lat* 'paper' and *kal*, meaning 'spear, pierce, poke, cut' inter alia (the association with 'poke' may be an Australian pattern, compare Gurindji *turrp-turrp-kaji* 'poke-RED-AGENT'). In addition, Comanche has *par*ʉ*boo?* /paa-t~~ʉ~~boor~~ʉ~~-ʔ/ 'water-write-NMLZ' for an 'ink pen' specifically, and Abzakh Adyghe *ṡḥerə-tχe* 'with.head/upperside-write.' In Hani, there is a semianalyzable term also denoting a 'writing brush' (a meaning also associated with 'pen' in Mandarin) with the identifiable constituent meaning 'to write' and 'to rot, decay' (the other one, *dul*, means 'similar, true to life' when occurring on its own.)

Sora *ido:l'kappa:n* contains *'kappa:-* 'wing' alongside *id'o:l-* 'write,' and Khoekhoe *xoa-ǀam-mi* is analyzable as 'scrape/write-feather-3SG.MASC.' Both point to a pattern particulary common in the Old World (as also evidenced by its frequency in Indo-European, Buck 1949: 1290): since quills were once commonly used as a writing instrument, Basque, Nivkh, and Kildin Saami colexify 'pen' with 'feather' and/or 'quill' (the Basque term *hegats*

also means ‘fin’ and ‘eaves’ and may be related diachronically to *hatz* ‘finger;’ by another term Basque also colexifies ‘snowflake’), while Efik has *ntañwet* /ntañ'-ñwet/ ‘feather-paper’ and Hawaiian *hulu kākau* ‘feather write.’

In turn, ‘paper’ (which potentially colexifies certain types of documents inter alia, cf. section 87) is of course by functional contiguity frequently associated with ‘pen,’ as already seen in the Yir Yoront and Efik terms mentioned above. ‘Paper’ and ‘pen’ are colexified in Yanomámi (similarly, Huambisa colexifies ‘pen’ with ‘booklet’). Furthermore, Bororo has *bapera atugo epa* ‘paper writing instrument,’ Biloxi *akŭtxyi´ o^{n}´ni ~ akŭtxyi´ o^{n}* ‘paper make,’ Carrier *testłes-tcen* ‘paper-stick,’ and Miskito *ulb-aia dusa* ‘write-INF stick’ (compare also Baruya *pikariyita* /pikarya-yita/ ‘carving-stick,’ colexification of ‘log, rafter 2x4, a limb, a young tree, a smaller log, any piece of wood that is like a stick’ with ‘pencil’ in Wintu, and the fact that the Bora term is derived from a verb meaning to ‘write’ by a classifier for small sticks). Rather than ‘stick,’ the second element is ‘rod’ in Piro and Imbabura Quechua; due to colexification with ‘tree, wood’ and ‘stick,’ this association is also present in Ngambay. Another complex term of the lexical type where one of the constituents is ‘write’ is Santiago Mexquititlan Otomí *xini t'ot'i*, containing *xi* ‘hair’ and *ot'i* ‘to write.’ White Hmong has *cwj-mem* ‘stick-ink,’ a derived term from a base meaning ‘ink’ is also found in Central Yup’ik (*ingeg-cuun* ‘ink-device.for’), and a term where one of the constituents is ‘ink’ is featured in Yay. Rotokas and Aguaruna colexify ‘pen’ with ‘typewriter’ (Aguaruna also with ‘booklet,’ ‘notebook’), and Bislama and Takia with ‘paint,’ in both cases due to collapse of English source words *pen* and *paint*.

Other associations include: Hausa *alkalami* also can refer to ‘a digit in arithmetic’ inter alia, Muna *koroka* denotes a fibre rib of the sugar palm which is used as a pencil, Khalkha has *yzyg*, which is also a variant of *yzyg ~ ysyg* ‘writing, letter.’ Welsh *pin* also means ‘bobbin,’ and Japanese *fude* is analyzable as /fumi-te/ ‘text-hand.’ Nivkh colexifies ‘pen’ and ‘pointed drill.’ Hawaiian *peni* (< Engl. *pen*) has fallen together with English *penny*, which was also borrowed, and Rotuman *pene* indigenously also means “to emit an odour.”

89. The Rope

Representation: 86%
Motivated: 55.5%
Thereof Analyzable: 5.1% Thereof Colexifying: 50.4%
Thereof by Contiguity: 5.1% Thereof by Similarity: 32.4%
Recurrent associated meanings: thread/string/cord/twine, vine/climbing plant, tie, line, lasso, strap, fishing line, badge of rank, cable, belt, bundle of rice, thong, whip, umbilical cord

The typical structure for many artifact terms - derived terms from the associated action that can be performed with the artefact - is relatively rarely found for this meaning. Muna *ka-tapu* is analyzable as ‘INSTR-tie/tether,’ and such derived terms are otherwise only found in Khalkha (where the derivative also means ‘hitching post’ and ‘training a horse for a race’), Chickasaw, Central Yup’ik, and Tehuelche (see Buck 1949: 550 for evidence from Lithuanian as well as from cognates meaning ‘cord, band’ and ‘bind’ respectively in San-

skrit and Greek). Kyaka, Khalkha, and Wintu terms directly colexifies 'rope' and 'to tie' (and there are sometimes redundant complex terms on the basis of the relevant root). More frequently, namely in 20 languages (Ngambay, Baruya, Buin, Kwoma, Kyaka, Mali, Rotokas, Sahu, Kosarek Yale, Sora, Wintu, Bora, Chayahuita, Hupda, Tsafiki, Wayampi, Fijian, Sedang, Takia, and Bislama), is colexification with (a specific kind of) 'vine' or 'climbing plant' generally, either by perceptual similarity or, more likely, by provenience contiguity (see Laycock 1970: 1160 for New Guinea specifically). Very frequent is colexification with 'thread,' 'string,' 'cord,' and/or 'twine,' found in Buli, Hausa, Katcha, Khoekhoe, Mbum, Ngambay, Rendille, Yoruba, Berik, Burarra, Gurindji, Kaluli, Kyaka, Lavukaleve, Muna, Rotokas, Kosarek Yale, Yir Yoront, Abzakh Adyghe (inter alia), Badaga, Basque, Bezhta, Chukchi, Khalkha, Kildin Saami, Sora, Cheyenne, Chickasaw, Ineseño Chumash, Haida, Kashaya, Lesser Antillean Creole French, Pawnee, Pipil, Quileute, Tuscarora, Wintu, Yuki, Central Yup'ik, San Lucas Quiaviní Zapotec, Copainalá Zoque, Arabela, Aymara, Bororo, Cashinahua, Cavineña, Hupda, Jarawara, Lengua, Piro, Tsafiki, Yanomámi, Bislama, Hani, Hawaiian, Bwe Karen, Manange, Samoan, Sedang, Takia, Tetun, and Yay (see Buck 1949: 550 for scattered Indo-European evidence). Somewhat similarly, Central Yup'ik has *qip'arpak* /qip'aq-rpak/ 'thick.hand-twisted.thread.large.' Many languages of Mesoamerica and adjacent areas colexify 'rope' with 'lasso.' This is the case in Itzaj, San Mateo del Mar Huave, Xicotepec de Juárez Totonac, Copainalá Zoque, and Cubeo. The relevant Xicotepec de Juárez Totonac term *tasiuj* is derived from *sihuí* 'to bend,' and the Wintu term contains a verb meaning 'to swing a rope or lasso.' In Rotokas, 'rope' is *iroiro*, which appears to be reduplicated from *iro* 'belt,' and similarly Bororo has *kogu-ia* 'belt-opening;' Kolyma Yukaghir colexifies 'belt' with 'rope.' Buli, Hausa (among other meanings), and Bislama colexify 'rope' with 'badge of rank' (Ngambay also with 'rank,' alongside 'trap'), Khoekhoe, Ngambay, Mali, Bezhta, and Cashinahua with 'fishing line' (Khoekhoe also with 'pore'), Carrier, Haida, Pawnee, Cubeo, Jarawara, and Hawaiian with 'line' more generally, Buli, Embera, and Hani with 'cable,' Pawnee and Hawaiian with 'thong,' and Kaluli, Chukchi, Cavineña, Jarawara, and Hawaiian with 'strap.' Muna *kalolai* 'rope for lowering things' is derived from *lolai* 'umbilical cord, to lower with a rope,' and Central Yup'ik colexifies 'rope' with 'umbilical cord.' Embera *hĩŋkará* also means 'whip, lash' when used with masculine gender and 'reed, cane' with neuter gender, and similarly, Hawaiian *kaula* also means 'whiplash,' alongside 'arc of circle' and "chain, as used by surveyors and engineers." Koyraboro Senni colexifies 'rope' with 'bundle of rice' inter alia, and Muna *kakoo* is derived from *koo* 'bundle of rice.'

Other associations include: Hausa *tuke*, meaning 'rope, thick string' in the dialect of Kano, otherwise means "twist together all of the material, e.g. as in making rope" inter alia. Ngambay *gól* also means 'to arrange, reconcile,' and Swahili *kamba* also 'shrimp, prawn.' Buin *kuuku* is also a female name. The Burarra term *murndurn* also means "group, work party, clan or tribe," and Kyaka *puu* also means 'bandage' among many other things, while *pungi* also means 'liver.' Ngaanytjarra *purturru* also denotes a 'hair string' (presumably used as a rope) and 'wool,' while One *apa* colexifies 'rope' with 'rattan.' Rotokas *koroviri* seems to be analyzable as 'fruit-twist.something,' and Sko *à* also means 'clear, shiny.' Toaripi *horou* also means 'intestines,' Kosarek Yale *heing* also 'eye,' and Badaga *agga ~ hagga*

also ‘plow’ and ‘connection.’ The Basque term *soka* is also used with the meanings ‘halyard’ and ‘dress,’ and Khalka *ujaɣasu* also means ‘knot,’ while *degesy(n)* is also an obsolete unit of measurement. Sora *lua:dən* also denotes “fibrous bark” as well as a specific type of string “with knots to indicate payments made to creditors.” Chickasaw colexifies ‘rope’ with ‘yarn’ and ‘leash,’ and Lesser Antillean Creole French with ‘chord, note’ (presumably due to phonological collapse of French *corde* and *accord*). Kiliwa *t+ha?-q=h-?+nyat-tay-u* ‘lead rope’ is analyzable as ‘OBJ+mouth-ABL/ALL=3-??+pull-FREQ-PL,’ Nuuchahnulth colexifies ‘rope’ with ‘in line,’ whereas in the Santo Domingo de Guzmán dialect of Pipil, *kwerda(h)* is also ‘a measure of land.’ Tuscarora *uhsì·reh* also means ‘wick’ and ‘wire,’ Wintu *ċe·k* also ‘ropelike root(s),’ and *ṗaλi* also ‘grapevine branches.’ Central Yup’ik *qecik* means ‘skin rind, scab’ and in the Norton Sound-Unaliq dialect also ‘rope.’ Copainalá Zoque colexifies ‘vein,’ and Bora *wáábya-u* is analyzable as ‘hammock-CL.round.’ Guaraní *sâ* also means ‘slavery,’ as well as, verbally, ‘to be fastened with a rope.’ Piro *tsa* also means ‘fiber,’ Wayampi *yã* also denotes a tree species, Fijian *dali* also ‘ten cuttle fish tied together,’ Bwe Karen *-bli* also ‘nest,’ Kapingamarangi *hali* also “to leak, to flow, to ooze,” Manange *1ts^{h}o* also ‘swamp,’ and Bislama *rop* also ‘tape of cassette.’

90. The Scissors

Representation: 71%
Motivated: 28.6%
Thereof Analyzable: 24.8% Thereof Colexifying: 4.3%
Thereof by Contiguity: 19.3% Thereof by Similarity: 7.9%
Recurrent associated meanings: cut, knife, tongs, cloth, razor, clip, pinch

Words for ‘scissors’ (or ‘shears’) are often derived from verbs meaning ‘to cut’ generally or more specific types of cutting, such as Blackfoot *sisóya’tsis* /sisayi-a’tsis/ ‘cut.into.strips-INSTR’ or Sedang *kơnep*, which is derived by the nominalizing infix *<ơn>* from *kep* ‘cut hair’ (see Buck 1949: 560 for the situation in Indo-European). Such terms are also attested in Chickasaw, Pawnee, Yaqui, Abipón, Ancash Quechua, Toba (where the term colexifies ‘knife’ and ‘razor,’ as is the case in Buli), Wayampi, Yanomámi, and Samoan, while there is a semianalyzable term in Bezhta. The Comanche, Kiowa, and Miskito terms feature an additional constituent meaning ‘cloth’ (e.g. Comanche *wana koo?* ‘cloth cutter;’ this term is archaic), and Muna colexifies the verbal reading ‘to cut’ (alongside “cross to the other side”) and the nominal one as ‘scissors’ directly. There are also languages which have a specific verb meaning ‘to cut with scissors’ (Biloxi, Arabela, Central Yup’ik), with the noun derived from it. Efik *ufañ'kpö* contains *fat* ‘clip, cut with scissors’ and *ñk'pö* ‘thing,’ and similarly, Fijian has *i koti* ‘DERIV clip/shear.’ In two sampled languages, Japanese and Central Yup’ik (Nunivak island dialect), terms are derived from a verb meaning ‘to pinch’ rather than ‘to cut’: *hasam-i* ‘pinch-NR’ (colexifying ‘scissors of lobster’) and *nunuutek* /nunur-(u)n/ ‘pinch-device.for’ (this term is formally dual). Similarly, the Arabela term *tuquetaja* is derived by instrument nominalization (*-taja*) from *tuquenu* ‘to pince bare.’ A semianalyzable term involving a verb meaning ‘to pinch’ is found in Quileute. Otherwise, it is frequent cross-linguistically to have complex terms for ‘scissors’ based on other arti-

facts with similar function and an additional element indicating the differentia specifica, which is in this case often the presence of the holes serving as handles for the fingers. The source artifact is most commonly 'knife': Sko *tanglílong* contains *tang,* which is a general term for blades and hence also can refer to a 'knife,' and *long* 'key, hole' (there is a word *lí* meaning 'cross-pole for roof or floor' inter alia), Cheyenne has *hótâxová-mota* 'crosscut-knife,' and Central Yup'ik *nuussicuak* /nuussig-cuar/ 'knife-little.one;' the Yanomámi term for 'scissors' is that for 'knife,' amended by the quantal classifier *-kɨ* (see § 4.4.1), and there is a semianalyzable term featuring a constituent meaning 'knife, sickle' in Oneida, and a redundant term with a constituent meaning 'knife' is present in Mandarin Chinese (note also Tundra Yukaghir *aŋńalď'oɣoje*, literally 'knife with a mouth,' Nikolaeva 2006: 106). Toba, as mentioned above, colexifies 'scissors' with 'knife' directly, and similarly, Buli *poning* (related to *poni* 'to shave, cut hair') also denotes a small knife for shaving. The cross-linguistic situation is thus exactly parallel to that within Indo-European: most frequent are terms derived from 'cut,' with the second most association being that with 'knife' (some Indo-European words for 'scissors,' notably in Celtic and Slavic, are formally the plural of or derived from words for 'knife,' Buck 1949: 560). Similarly, the handles are also used as the conceptualization source in Rotokas, which has *kaporoto*, containing *kaporo* 'space between objects, (mountain) passage.' Ket has a term making reference to both 'knife' as well as another perceptually similar artifact, 'tongs': *atəpəl doˀn* /atəp-ul doˀn/ 'tongs-handle knife.' An association with 'tongs' or 'pliers' is also realized by the complex Kapingamarangi term *di kabi-kabi* 'ART fire.tongs-RED;' *kabi* also has the additional meaning 'to hold something between two other things.' Furthermore, 'scissors' and 'tongs' are colexified in Yir Yoront, Khalkha, and Hawaiian (here, also with other implements similar in function).

Other associations include: Yir Yoront *thaminhwaw* contains *minh* 'animal,' and Badaga *katri koːlu* contains *koːlu*, meaning inter alia 'stick' and 'skewer.' Chukchi *weŋətkuneŋ* contains *weŋ* 'yarn.' Kildin Saami *rūvv't* also means 'iron' and 'trap.' Haida has a term based on a verb meaning "make go (apart) into two pieces" prefixed with a verbal classifier for tongs or scissors. Nez Perce *capáˑkaƙiwkaʔs* is analyzable as /cepéˑ-kéˑ-ƙíw-ʔs/ 'by.pressure-with.teeth-take-??,' and Nuuchahnulth *ƛ̓ap-ẏak* as 'straddle-tool.' As a verb, Wintu *ṗin* also means 'to squeeze' as of long objects, as well as 'scissors-like leg movements.' Bora *majchówa* appears to be derived from *majcho* 'food, eat' (there is another term featuring an element meaning 'house' and 'triangular frame'); perhaps there are errors in lemmatization in the consulted source. Cavineña *tishira* is also a term for wood that sustains the roof of houses (accidentally, if *tishira* < Span. *tijera*), and Hupda *hăy'b'ah* is analyzable as 'shear-flat.thing.' Tehuelche *ʔepernwe* is derived from *ʔepˈer ~ ʔepˈere* 'to crop,' and there is a semianalyzable term featuring a constituent with this meaning in Toba. Bislama *sisis* (< Engl. *scissors*) also means 'close friend' and, verbally, "to stick closely to someone, to hug sexual partner," while Vietnamese *kéo* also means 'to pull.'

91. The Shoe

Representation: 81%
Motivated: 29.4%
Thereof Analyzable: 20% Thereof Colexifying: 9.4%
Thereof by Contiguity: 21.0% Thereof by Similarity: 1.1%
Recurrent associated meanings: foot, put on/wear, sandal, footprint, moccasin, skin, walk

Most common among the lexico-semantic associations for 'shoe' ('boot,' 'footwear') is that with 'foot' (found also by derivation in Ancient Greek and by compounding in Welsh, Buck 1949: 428). Among terms betraying this association, a particularly frequent subtype is constituted by complex terms where the second constituent is a verb meaning 'to put on' or 'wear,' such as Kaluli *gib-a sa:ga:la:sen* 'foot-?? put.on.' Such terms are also found in Laz, Haida, Yuki, Miskito, and Bwe Karen; similarly, Samoan has *se'e-vae* 'slide/slip-foot/leg.' In Australia, it is particularly common to colexify 'shoe' with 'foot' directly. This is found in Gurindji, Ngaanyatjarra, Yir Yoront, dialectally in Basque (which otherwise has a derived term), as well as by a prefix in Comanche (Ngaanyatjarra, Yir Yoront, and Comanche also colexify 'footprint,' and Comanche also 'trail,' while Yir Yoront has an additional complex term containing elements meaning 'human' and 'go in'). Otherwise, the secondary associations are manifold. Efik has *ik'pa-uküt'* 'leather-foot/leg,' Ngambay *né-gɔ̀l* 'thing-foot/footprint,' Pawnee *asuuruʔ*, analyzable as /as-hur-uʔ/ 'foot-place-NOM' (originally denoting the 'moccassin' specifically, now 'shoe' generally;' this type of autohyponymy is synchronically still present in Cheyenne, Nez Perce, and Yana, while in Buli, Hausa, Rendille, Sko, Toaripi, Hani, Hawaiian, Rotuman, and Samoan, relevant terms also mean 'sandal'), Bora *túhapáájɨ* /túhaá-pa:hɨ/ 'foot-SCM.hole,' Bororo *bure tadawu* 'foot which.is.under' (note that Greek and Sanskrit terms for 'shoe' is derived from a verb meaning 'to bind under,' 'to tie under,' Buck 1949: 428), Cashinahua *bin tae* 'rubber foot,' Guaraní *py-ao* 'foot-clothes,' Kaingang *pẽn né* 'foot container,' Maxakalí *pata-xax* 'foot-cover/skin/bark' (the Tuscarora term, non-transparent today, might have been made op of elements meaning 'foot' and 'cover' originally, and a compound featuring elements with these meanings is attested in Persian and is etymologically recoverable for Welsh, Buck 1949: 428), Yanomámi *mamikɨtitioma* contains *mami* 'foot' and *titiha-ɨ ~ titihi-aɨ* 'put in,' and Kapingamarangi has *hii wae* 'package/wrap leg/foot.' The Mali term *alēcharachi* is derived from *lēchar* 'foot,' and Chukchi *jeɣət* is grammatically the plural of the word for 'foot.' Furthermore, Kyaka has *kimbu suu*, with *kimbu* meaning 'foot, leg' and *suu* being a loanword from Tok Pisin, that is, however, also a native lexical item meaning 'drain, trench' inter alia. Semianalyzable terms where the identifiable constituent is 'foot' or 'foot, leg' are found in Kemtuik, Sentani, Sko, Carrier, Upper Chehalis, Wappo, and Guaraní. Less frequent associations are that with 'skin' (which is one of the meanings colexified in one of the constituents of the Maxakalí term mentioned above), found in Meyah, which has *mek mei ofos* 'pig coastal skin,' and occuring in Ngambay by colexification, and that with 'to walk' in Abipón (*n-acaR-haR-late* 'POSS.INDEF/3SG-walk-the.one.who-LOC;' this term colexifies 'ladder'). Badaga colexifies 'walk on, step on' and other meanings with 'shoe.' Furthermore, Yana *niɬiiwau(na)* might contain the root *ni-*, meaning 'for a male to walk.'

Other unique associations include: Hausa *takalmi* is also used metaphorically with the meaning 'provisions for a journey,' Muna *kolo* 'wooden shoe' (in this sense a loanword from Bahasa Indonesia ultimately going back to Dutch *klomp*) also indigenously means 'sour' and 'to carry someone on the back,' and Ngaanyatjarra *tjina* may, alongside 'foot,' 'footprint,' and 'shoe,' also refer to 'claws, talons,' 'tracks' and means 'on foot' adverbially. Rotokas *kuroea* is also the name of a species of vine with "leathery appearance," and Basque *zapata* is also used with the meanings 'threshold,' 'chassis,' 'frame,' 'buttress' and "thin wooden fence." Bezhta *halałco* is made up of the word for 'leg' and the essive case marker. The San Mateo del Mar Huave term *socol napiüc* appears to contain *socol* 'corner,' and Nez Perce colexifies 'shoe' with 'horseshoe.' Cubeo *cɯraido* consists of *cɯrai* 'ground' and *-do*, the classifier for hole-like objects. Guaraní colexifies 'shoe' with 'sock,' Fijian *vāvā* also denotes the rungs of a ladder, Hani *seiqnaov* 'shoes, sandals' might be related to *seiq*, meaning inter alia 'hoof' and 'bamboo stick on which to roll cotton in preparation for spinning it,' Hawaiian *kā-ma'a* is analyzable as 'CAUS-bind,' Takia *su* also means 'breast, udder, milk' (due to collapse of an inherited term with Engl. *shoe*), and Yay colexifies 'shoe' with 'to put out of the mouth.'

92. The Street

Representation: 89%
Motivated: 26.2%
Thereof Analyzable: 9.4% Thereof Colexifying: 16.0%
Thereof by Contiguity: 9.9% Thereof by Similarity: 6.2%
Recurrent associated meanings: manner/method/system, town/settlement, door, journey, go/walk, market, footprint, between, fare/freight, waterway, Milky Way

The most frequent association for this meaning (additional glosses such as 'way,' 'trail' etc. are disregarded in the following discussion) is a metaphorical abstraction, namely to 'manner,' 'method,' or 'system.' Eleven languages in the sample, Koyraboro Senni, Ngambay, Rendille (where the term also means "right thing to do," "good way (of behaving)"), Kwoma, Basque, Greek, Guaraní, Rotuman, Samoan, Tetun, and Bislama feature this pattern of colexification. Similarly, the Burarra term is also used with reference to the 'way of living,' and the Kyaka term also means 'category.' In two areas of the world, New Guinea and the American Northwest, 'street' is colexified with 'door' and/or 'doorway' in some languages. This pattern is attested in Dadibi, Kaluli, Upper Chehalis, Haida, and Nuuchahnulth. Three sampled languages, Efik, Chickasaw, and Fijian have complex terms for street where one of the constituents has a meaning akin to 'between,' for instance Chickasaw has *okla-ittintakla'* 'town-between' and Fijian *saqata ni koro* 'interval POSS village.' Indeed, associations with 'town' or 'settlement' are themselves relatively frequent. Badaga, Basque, and Ancash Quechua colexify these meanings (and Kildin Saami colexifies 'street' with 'place in town'), Tuscarora has *yuta?nakáhrę·?,* analyzable as /yu-ta'n- kahrę(w)-·?/ '3SG.NEUT.PATIENT-settlement-be.an.opening-STAT' (compare Muna *kabhongka*, derived from *bhonkga* 'to crack, smash, for a road to open'). The relevant Buin term also is the name of a particular village. In four sampled languages, Chukchi, Abipón, Aguaruna, and Guaraní,

words for ‘street’ derived from verbs meaning ‘to go’ or ‘to walk’ are found (evidenced diachronically in Baltic, Buck 1949: 721); the relevant Guaraní term also means ‘pedestrian’ and ‘concubine.’ As an example of such a derived term, Chukchi *təlan* /təle-n/ ‘go-LOC’ may serve; note also that the Bora term *ullé*j́*íjto* is derived from *ullé*j*e* ‘to travel’ by means of the classifier *-j́íjto* for lines or roads. Moreover, Rotuman colexifies ‘street’ and ‘to go, come’ directly.

Khoekhoe and Piro colexify ‘street’ with ‘footprint,’ and similarly, Central Yup’ik has *tumyarak* /tuma^{e}-yarak/ ‘footprint/track/trail-device.for.’ Buli, Ngambay, Basque, Greek, and Rotuman colexify ‘street’ with ‘journey,’ while Khalkha, Hani, and Mandarin colexify it with ‘market’ (and similarly, Yoruba with “public thoroughfare”), and Wayampi (alongside other meanings) and Hawaiian with ‘waterway’ (for which compare the association between ‘way, trail’ with ‘river’ reported in section 47). Similarly, Hausa colexifies ‘channel.’ Finally, Rotuman and Bislama colexify ‘street’ with ‘fare, freight,’ and Hausa and perhaps Tasmanian (Plomley 1976: 408) colexify ‘Milky Way’ (Hausa also ‘beehive’ inter alia).

Other associations include: Buli *siuk* also means ‘permission, right’ inter alia, and Efik *añ'wa*, denoting a ‘principal street,’ is analyzable as /a-ñwañ'a/ ‘NMLZ-widen’ (compare the derivation of a word for ‘street’ from ‘wide’ in Ancient Greek, Buck 1949: 720). Swahili *barabara* also means “proper, as it should be,” Buin colexifies ‘street’ with ‘access,’ and another Buin term, *rootu*, is the outcome of borrowing of both Tok Pisin *lotu*, meaning “church, worship, church service” and English *road*, and consequently has both readings. Kwoma *nobo* also means “illegitimate; born out of wedlock,” Lavukaleve *lake* also means ‘fire,’ and Muna *sala* is also used with the meanings ‘sort’ and ‘thing,’ alongside ‘trousers’ and other meanings. Toaripi *oti-haro* is analyzable as ‘place-head/chief,’ Rotokas colexifies ‘road, path, way’ with ‘line,’ and Yir Yoront *yalq* also denotes groups of animate beings, such as a flock of animals or school of fish. Badaga *ke:ri* also denotes the ‘frontyard’ and “work-space in front of houses” and is “an appropriate measure of land area.” Basque *bide* also can refer to the ‘platform’ at a train station inter alia, *kale* also to the ‘eye of a needle,’ again next to other meanings. Greek *drómos* can also refer to a ‘distance,’ as well as ‘speed’ and a ‘race.’ Japanese *tōr-i* is analyzable as ‘pass-NR.’ Khalkha *ɣudumzi(n) ~ ɣudamzi* seems to be related to *ɣudum* “passage, thoroughfare, hallway, corridor,” *zegeli* is identical segmentally to one of the variants of *zegeli ~ zegele* ‘debt, loan,’ and Sora *'taŋgo:rən ~ taŋo:rən* also means ‘occasion.’ Biloxi *nětkohi´ ~ natkohi ~ nĭtkohi ~ nŭtkuhi ~ nŭtkohi* contains *nĕ* ‘to stand,’ and *kohi ~ ḳụhi´ ~ kụ´hi ~ ḳuhi* ‘up, high.’ Comanche *kawonokatu̲* is said to literally mean “wolf separate camp.” Haida colexifies ‘street’ with ‘ladder,’ and Kashaya *hiʔda* contains *ʔda* ‘extend, stretch.’ Aguaruna *jínta* also means ‘faculty, specialty’ and ‘terrain,’ Bora *huúva* might contain *huú* ‘tube,’ Bororo colexifies ‘railway’ and ‘trace of cobra,’ and Chayahuita *pa'tërinso'* is derived from *pa'tërin* ‘to weave.’ Guaraní *tape* colexifies ‘street’ with ‘art’ and ‘religion,’ while the Miskito term *yabal* also means ‘mouth, tongue, way of speaking’ and *bila* also ‘mouth, opening,’ ‘center, inner part, space’ as well as ‘word’ and ‘language.’ Fijian *sala* also inter alia means ‘to climb or creep upon,’ and Bwe Karen *klɛ* also ‘to fix on’ inter alia. Kapingamarangi *ala* also means ‘responsibility’ inter alia, and Lenakel *suatu* also ‘course of ship’ and the “traditional exchange links between individuals or de-

scent groups." Malagasy *arabè* is analyzable as */àraka-bè/* 'following-big/much.' Rotuman colexifies 'meaning,' 'condition,' and other things. Sedang *tróang* also means 'sentence,' Vietnamese *đường* also 'sugar,' and Yay *ran*^{l} also 'to see.'

93. The Table

Representation: 77%
Motivated: 35.5%
Thereof Analyzable: 23.6% Thereof Colexifying: 11.9%
Thereof by Contiguity: 19.5% Thereof by Similarity: 7.9%
Recurrent associated meanings: eat, furniture, food, to place on, board, floor, flat, dish, tree/wood, place, restaurant

Terms for 'table' derived from verbs meaning 'to eat,' such as Carrier *uḳwez-eṭał* 'on.it-eat' are found in Chickasaw, Kiliwa, Lake Miwok (where the term colexifies 'food, groceries'), Nez Perce (where the term colexifies 'restaurant, inn' and 'kitchen,' the former pattern is also attested in the Norton Sound -Upaliq dialect of Central Yup'ik), Wintu, Central Yup'ik (again, only in the Norton Sound - Upaliq dialect), Abipón, Arabela (where the term is derived by an agentive nominalizer and the resulting term hence can also refer to an 'eater, one who eats,' as is the case in Guaraní), Guaraní (where the term also means 'refectory,' 'dinner guest,' and 'tablecloth'), and Yanomámi. A semianalyzable term is present in Upper Chehalis, and similar complex terms of the lexical type are found in Biloxi, where *a´diṭo*^{n}*ni ~ adito*^{n}*´* contains *ti* 'to eat' and *o*^{n} *~ o*^{n}*ni* 'to make, do,' Blackfoot (*iitáísooyo'p* /iit-á-iso-ooyi-o'p/ 'LOC-DUR-on.a.horizontal.surface-eat-21.NOM'), Cheyenne (*táxemésêhestôtse* /táxe-mésehe-hestôtse/ 'upon-eat-thing'), Haida (*ga taa daan* 'INDEF.PRONOUN eat place), Kiowa (*p´į-á·* 'eat-board'), and Yana (*moorił?mauna* /ma-ri?mau(na)/ 'eat-place'). Moreover, Yir Yoront has *yo-penpn may pay+n* 'wood-flat food eat+NOUN.THEME.FORMATIVE' for 'dining table' specifically, and indeed, complex terms based on 'food' rather than 'eat' are also relatively frequent (and terms betraying this connection dominate in Indo-European, Buck 1949: 483). Thus, Kyaka has *nenge rate* 'food bench/shelf,' Kashaya *ʔacaʔ ʔama bumucid=tol* 'people food eat=on,' and Wintu *ba·s top-i* 'food used.for-NOMINAL.STEM.FORMANT,' but a particularly frequent combination, especially in North America, is with verbs meaning 'to put,' 'to place on,' as in Oneida *atekhwahlákhwaʔ*, which is analyzable as /ate-khw-hel-hkw-aʔ/ 'SRFLX-food-set.on.top.of/place.on-INSTR-HAB.' Such terms are also featured in Tuscarora and Yuki, and, somewhat similarly, Pawnee has *rakaraaraaruukitaʔiituʔ* /rakaraa-raar-huukita-iit-uʔ/ 'dishes-place-on.top.of-in.a.line-NOM,' Hawaiian *pā-kau* 'dish-put,' and *papa kau-kau* 'flat.surface RED-put' (though note that *papa*, which also has still other meanings, can also refer to a 'table' itself'), Cashinahua *tsaun-ti* 'put-INSTR,' and Yuki also has the alternative term *hąwąy ṭu·k-uł* 'food dish-INSTR.' There is a semianalyzable term where the identifiable constituent means 'food' in Huambisa, and the association is present also in Abipón due to colexification of verbal 'eat' and nominal 'food.' In fact, in Lake Miwok and Yanomámi, which feature derived terms from 'eat,' 'table' is itself colexified directly with 'food.' In Cubeo, *tʉoiva* is derived from *tʉoyʉ* 'to serve food' by means of the classifier *-va* for broad and flat objects, while Ket *laʔm* also itself means 'flat' as an adjective. Alongside

this and the Yir Yoront term already mentioned above, an association between 'table' and 'flat surface,' among other meanings, is found by colexification in Hawaiian and possibly in Samoan, where *laulau* is 'table' and *lau* both 'leaf' as well as a classifier for flat or thin objects. A similar association to that in Yir Yoront and Kiowa, where the relevant terms contain constituents meaning 'wood' and 'board,' is found in Bororo, which has *irá* /i-ra/ 'tree-bone.' Baruya, Dadibi, Toaripi, and Hawaiian colexify 'table' with 'floor,' and finally, it is not uncommon for 'table' to be colexified with other pieces of furniture, which happens in Noni, Baruya, Burarra, One, Rotokas (by the term *koara ua*, putatively analyzable as 'put.together NARROW.OBJECT'), Toaripi, Basque, Ket, Khalkha, Welsh, Wintu, Bororo, Cashinahua, Embera (also with 'saddle'), Hawaiian, and White Hmong, in which case the common semantic denominator seems to be that pieces of furniture are typically human-made raised surfaces. Similarly, Ket, Welsh, Ineseño Chumash, and Bororo colexify 'table' (also) with 'board.'

Other associations are: One *simpa* also means 'bridge,' and Khalkha *sirege(n)* colexifies 'table' with 'throne' as well as 'altar' and 'feast, banquet.' Welsh *tabl* also means 'tablet.' Nez Perce *tíˑmewesˑ* is analyzable as /tíˑmˌe-nweˑs/ 'write-LOC,' and Wintu *panti?ilestopi* 'table, desk' contains *panti* 'on top, on' and *top* 'used for.' Guaraní *arikapa* is a neologism containing *ári* 'on top' and *pa* 'everything.' Hani *lolbieil* contains *bieil*, meaning 'wide' and also acting as a classifier for 'chairs' and 'tables.' Hawaiian *papa* means 'flat surface' generally, and therefore colexifies 'table' with notions such as 'reef,' 'layer,' and others, while *pūneʻe*, meaning also 'movable couch' and rarely 'pew,' appears to be related to *neʻe*, which means 'moving along little by little or by fits and starts, to step' inter alia. The Sedang term *kơˈbang* also means 'blackboard,' and Yay *soŋ*^{4} means 'to scrape' in verbal usage.

94. The Toilet

Representation: 50%
Motivated: 54.4%
Thereof Analyzable: 49.3% Thereof Colexifying: 6.8%
Thereof by Contiguity: 46.1% Thereof by Similarity: 2.4%
Recurrent associated meanings: faeces, house, bathroom, place, bush/forest, hole, small, bathe/bath, outside, room, relieve, water, wash

The most common association for 'toilet' (or 'latrine') is by functional contiguity with 'faeces' or 'defecate.' Within terms betraying that association, the most dominant subpattern are terms of the lexical type with 'house' being the additional constituent, as in Dongolese Nubian *úññíŋkã* /úññɪ-ŋ-kã́/ 'excrement-GEN-house.' Such terms are also found in Anggor, Baruya, Dadibi, Kaluli, Kwoma, Kyaka (with an additional constituent meaning 'origin'), Rotokas, Comanche, Kashaya, Santiago Mexquititlan Otomí, Pawnee (with an additional constituent meaning 'to be going'), Wappo (which also has *mahkúyeˑma čhùya*, containing an element meaning 'exit' alongside 'house'), and Yay, some of them denoting an 'outdoor toilet' specifically. Note also that the Bora term *naméja* is derived from *name* 'excremement' by the classifier *-ja* for houses, as well as Kiowa *sʜ'tsoue-tou* 'urine-house'

for 'urinal.' Otherwise, 'place' is the meaning of the second constituent in Kwoma *she pii eem* 'faeces defecate place,' Japanese *ben-jo* 'faeces-place,' and Ket *horoks baˀŋ* /ho'q-d-oks baˀŋ/ 'excrement-POSS-stick place,' Hawaiian *wahi hoʻo-pau pilikia* 'place CAUS-finish trouble' (which also has *kiʻo-na* 'excrete-NMLZ,' colexifying 'dung heap, dump'), while Greek directly colexifies 'place' and 'toilet' (as well as 'role' and 'party'). Other terms of the lexical type with 'faeces' are Ngambay *buwá-sìn* 'hole-faeces' (compare Kyaka *yuuwuali anda ~ iwali anda* 'deep.narrow.hole/pit house,' and the colexification of 'hole' with 'toilet' in Hawaiian, among other meanings), Hupda *y'ӕ-ʔteg* 'faeces-thing,' and Hani *xiqduq*, with *xiq* meaning 'faeces' and *duq* meaning 'to dig,' but also functioning as a classifier for pools of water and the sound of drums or thunder. Derived terms also occur, namely in Central Yup'ik, Aguaruna, Bora (as mentioned above), Wichí, and perhaps in the Cuisnahuat dialect of Pipil; semianalyzable terms are found in Chukchi, Nez Perce, and Guaraní. Similarly, Yoruba *ibi-ìgbọ̀nsẹ̀* is analyzable as 'place-relieve,' and Gurindji has *warlp-kaji* 'relieve.self-AGENT.'

Complex terms where one of the constituents means 'house' are also found with second constituents other than 'faeces:' Bislama has *smol-haos* 'small-house,' which is paralleled in Khoekhoe and Samoan, Abzakh Adyghe has *psə-wəne* 'water-house' (for the association with water, compare also Koyraboro Senni *hari-mun-doo* 'water-pour-place'), Welsh *tŷ bach* 'corner/hook house,' Cheyenne *o'êhné-mâhéó'o* 'eliminate-house' and *too'hamé-mâheo'o* 'bathe house,' Santiago Mexquititlan Otomí analogously *nguu nsaha* 'house bath' (Carib moreover has a derived term from 'bath'), Samoan *fale-ʻese* 'house-other,' *fale-ui* 'house-go.along,' and *fale-vao* 'house-bush/forest' (note also the colexification of 'toilet' with 'bush/weed' or 'forest' in Ngambay and Lenakel), and Yay *raan*^{4} *pay*^{1} *rok*^{5} 'house go outside' for an 'outdoor toilet' specifically (note also the colexification of 'toilet' with 'outside' in Aguaruna); further, Hausa *ban'daki* is a contraction of *bayan daki* 'back house.' Semianalyzable terms where one of the constituents is 'house' are found in Rotuman and Samoan. Japanese *keshō-shitsu* is questionably analyzable as 'mascara-room,' and Vietnamese *phòng vệ sinh* as 'room protect life.'

Moreover, terms involving a verb meaning 'to wash' are found in Japanese and Bororo. Finally, 'toilet' and 'bathroom' more generally are colexified in Buli, Khoekhoe, Swahili (alongside 'prayer mat'), Cheyenne, Kashaya, Aguaruna, Bororo, Hawaiian, and Tetun.

Other associations include: Hausa colexifies 'toilet' and 'cesspit,' and another term also denotes the "[d]oubling of a consonant, and the sign over a written consonant to indicate the doubling" as well as "a kind of European-made silk material." Swahili *choo* is related to *oga* 'to wash body.' Chickasaw *aaombiniili' ~ aambiniili'* is analyzable as /aa-on-biniili-'/ 'LOC-APPL-sit.down-NMLZ' and also means 'chair.' San Lucas Quiaviní Zapotec *ta's x:tèe'n ba'nny* is analyzable as 'bowl of bathroom,' and Toba *nogoxoqui* is related to *alogo* 'clothes.' Rotuman *fạʻu* inter alia also means 'back' (note also that Buli *banjiri* borrowed from Hausa *bayan gida* 'behind the house').

95. The Train

Representation: 41%
Motivated: 52.2%
Thereof Analyzable: 45.6% Thereof Colexifying: 6.7%
Thereof by Contiguity: 10.0% Thereof by Similarity: 38.9%
Recurrent associated meanings: wagon/vehicle/car, fire, iron/metal/steel, boat/ship, smoke, land/ground, run, road/path, carry

It is common for motivated terms for 'train' to be of the lexical type, with one of the constituents meaning 'wagon,' 'car,' or 'vehicle' generally. The semantics of the second constituent varies considerably, but not without limits. Terms with 'fire' or 'fiery' being the second constituent are found in Abzakh Adyghe (*meś$^{ə?}$ek^{ə}* /maś$^{ə?}$e-k^{ə}ə/ 'fire wagon'), Khalkha, Kashaya, Lake Miwok, Hawaiian, Mandarin, and Vietnamese; note also Nez Perce *ʔalahíˑn* /ʔáˑla-híˑn/ 'fire-thing.with' and Yay *rua^{4} fi^{4}* 'boat fire.' There are also languages where 'smoke' or 'steam' is the meaning of the second constituent, as in Swahili *gari (la) moshi* 'car (of) smoke,' and also in Japanese (for a 'steam locomotive' specifically), Blackfoot and Chayahuita; note also that Wichí *wapulh* is a loanword from Spanish *vapor* 'steam.' Kanuri has *mààrá cídí-bè* 'vehicle ground-of.' Two Oceanic languages, spoken on small islands, showcase, similar to the situation in Yay just mentioned, a metaphorical transfer from 'ship' to 'train,' with the differentia specifica indicated by a term for 'land' or 'ground': Rotuman *'ahại la' ufa* contains *'ahại* 'ship' and *ufa* 'land,' and Fijian *sitima ni vanua* is analyzable as 'steamer POSS land,' while Hausa colexifies 'train' with 'boat' (and 'through,' see section 76 on this pattern). White Hmong has a complex term where 'iron' is one of the other constituents alongside 'wagon': *tsheb nqaj hlau* 'vehicle rail iron;' other complex terms on the basis of 'iron' are Cheyenne *ma'aataemeo'o* /ma'aata-meo'o/ 'iron-road,' Yuki *lil haʔ-oł ~ lil ham-oł* 'iron carry-INSTR' (compare Kiliwa *wa?=t-kw+lkwii-y-tay* as 'house=OBJ-WH+carry-ATT-FREQ' with the literal meaning "back-packer house" according to the source), and Malagasy *fiarandalamb'y* /fiara-n-làlana-vy/ 'vehicle-GEN-road-iron.' Santiago Mexquititlan Otomí colexifies 'iron' with 'car,' and 'train' is *ma-bo̲jä* 'long-iron/car' or *nju̲nu̲bo̲jä* /n-ju̲ni-bo̲jä/ 'NMLZ-to.tie-iron/car.' Biloxi has *yaduxta$^{n´}$ ta^{n}hi$^{n´}$* 'wagon run' and a verb meaning 'to run' also figures in the Chickasaw term *(itti') chanaa malili-'* 'wood roll run-NMLZ.' Other complex terms with 'vehicle,' 'car,' or 'wagon' are Khoekhoe *ǂnū-kuni-s* 'black-wagon-3SG.FEM,' Japanese *den-sha* 'electric-car,' and Oneida *tsyoʔslehtáˑkat*, analyzable as /s-yo-ʔsleht-ʔkaht/ 'REPETITIVE-NEUT.PATIENT-vehicle-move.fast.' Moreover, a semianalyzable term with 'wagon' is found in Blackfoot and with 'bullock cart' in Bwe Karen, while Comanche, Pawnee, and Wintu colexify 'train' with 'car' directly.

Other associations include: Buli *girigiri* is also an onomatopoetic word that "imitates a low rumbling noise" generally. Khoekhoe *ǁnubutas* contains the verb *ǁnubu* 'to shake, agitate,' and Gurindji *turrkalangarna* contains *ngarna* 'denizen.' The morphological analysis of Carrier *yenḳenekhĕs łênedîzṭî* remains unclear, but the "literal" translation offered by the source for *yenḳenekhĕs* is "land on it (being heavy) moves on," and *łênedîzṭî* contains elements meaning "attached together" (*łê*), "in a line" (*ṭî*) and "being several" (*ne*). Kiowa *'ą̆nk'ị̆ɲ-gɑ* is analyzable as 'go.along-NOUN.POSTFIX,' while Lesser Antillean Creole

French *twen* (< Engl. *tren* 'train') also means 'attendants' as well as 'noise, trouble.' Samoan *nofoa afi* is analyzable as 'sit-SUFFIX (fire/engine)' and colexifies 'chair, seat' and 'saddle.'

96. The Weapon

Representation: 51%
Motivated: 50.9%
Thereof Analyzable: 38.7% Thereof Colexifying: 14.2%
Thereof by Contiguity: 34.7% Thereof by Similarity: 0.9%
Recurrent associated meanings: tool, thing, rifle/gun, war, fight, sharp, army, bow, wound/to wound, defend, hurt, kill, throw, soldier, use, battle

Terms for 'weapon' are sometimes derived from verbs meaning 'to fight,' as in Xicotepec de Juárez Totonac *līta'lān* /lī-ta'lá-n/ 'INSTR-fight-INSTR' (and also in Sanskrit, Buck 1949: 1384). Such terms are also found in Kashaya, Miskito, and Malagasy, while Rotokas *upo purapara* contains *upo* "strike, slap, fight, murder" and *pura* "make, do, create," Yoruba has *ohun-ìjà* 'thing-fight,' Samoan *mea-tau* 'thing-fight,' and there is a semianalyzable term in Blackfoot. There are also terms based on verbs with similar but different semantics. Chickasaw has *ishhottopachi'* /isht hottopa-chi-'/ 'with hurt-CAUS-NMLZ,' Hawaiian *mea hōʻeha* 'thing hurt,' and a semianalyzable term containing a verb meaning 'to aim, hurt' is found in Yanomámi. The Carib term *owotopo* is derived from *owo* 'kill,' and Hawaiian has *mea pepehi kanaka* 'thing beat/kill man.' Laz *o-ťoç-aşe* is analyzable as 'DERIV-throw/shoot-DERIV' (similar derivation is found in Sanskrit for a term probably originally denoting missile weapons, Buck 1949: 1384), and Hupda *d'ap-b'ŭy-teg* perhaps as 'flesh-throw-thing.' Hausa colexifies 'weapon' with 'wound inflicted by a metal weapon,' and Nuuchahnulth *saʔa-čakʷ* is analyzable as 'wound/strike.with.weapon-tool.' More frequent than any of these associations, however, are terms where one of the constituents is 'war,' as in Kildin Saami *tuarr-viešš* 'war-thing.' Precisely parallel terms are found in Efik, Katcha, Noni, Yoruba (where 'war' is colexified with 'army' and 'battle' inter alia), Hawaiian (where the same remarks as for Yoruba apply), and Takia. Similarly, Carrier has *pê-netsepah* 'where-with-to.war' (alongside *pê-neznîł·u* 'wherewith-to.be.cuirassed'). A literal translation of Bora *ɨ́ɨneri metéhmémeíne* is "that with which we guard (defend) ourselves," and a Toba term for 'weapon' is derived from a verb meaning 'to defend.' A frequent pattern of colexification is that with (metal) 'tool,' found in Yoruba (by the analyzable term *ohun-i-lo* 'thing-NMLZ-use,' for which compare Tuscarora *yéčthaʔ*, which is based on the root *-ačT-* 'to use'), Badaga (where the relevant term also means 'door-bolt,' among other specific tools), Khalkha, Sora, Welsh, Highland Chontal, Lesser Antillean Creole French, and Central Yup'ik; this pattern is also common in Indo-European (Buck 1949: 1383-1384). Analogously, Japanese has *hei-ki* 'soldier-tool' (for the association with 'soldier,' note also Khalkha *cerig yn zemseg* containing *cerig* 'warrior, soldier, army' and *zemseg* 'ornaments,' Mandarin *wu3-qi4* 'military/valiant-utensil/appliance/apparatus,' and Vietnamese *vũ khí* 'martial tool.' Similarly, Guaraní *rairô rembiporu* is analyzable as /rairo tembiporu/ 'very.sharp instrument,' and San Mateo del Mar Huave colexifies 'weapon' with 'iron' and 'prison.' Similarly to the situation in Guaraní, Toaripi has *mare etau* 'sharp thing,' and Cheyenne *heškóvanëö'o*

contains *heškóv* 'sharp' and *ôhëö'o* 'thing;' the tentative literal translation offered is "sharp-by.hand(.thing)" and is marked as unsure in the source. Still more frequent, however, is colexification with a particular type of 'weapon' (autohyponymy). This is found with 'rifle' or 'gun' in Greek, Highland Chontal, San Mateo del Mar Huave, Xicotepec de Juárez Totonac, Yaqui, and Miskito (there are also two languages with complex terms on the basis of 'rifle': Chayahuita has *irapa inapita pochin ninin-so'* 'rifle these like do/be-3SG.SUB' and White Hmong the dvandva compound *riam phom* 'knife gun'), with 'bow' in Kyaka and Bororo, with 'sword' in Badaga, with 'slingshot' in Cheyenne, with 'machete' and, by a different term, with 'lance,' in Chayahuita.

Other associations include: Hausa *makami* is derived from *kama* 'to catch,' and indeed can also refer to a 'catcher.' The Muna term *ewanga ~ iwanga* contains *ewa* 'martial arts, to attack, oppose,' while *sandata* also means 'to supply, prepare' and 'supplies' (the meaning 'weapon' is due to borrowing of Bahasa Indonesia *senjata*). Welsh *erfyn* also means 'to beg, to pray,' and the Abzakh Adyghe terms contains the word for 'hand,' as does the Haida term. Kiowa *mɔ̀n-pà-tò·* is analyzable as 'hand-against-hold.' Bora *lliiñája* 'hunting weapons' might contain *lliíñe* 'punishment' and may indeed also refer to 'hunting,' there is a further derived term containing an element with that meaning. Santiago Mexquititlan Otomí *bojä* also means 'money,' Bororo *boe eiga* appears to contain *boe* 'thing' and perhaps *iga* 'splint.' Cayapa colexifies 'shadow, soul, reflection' (presumably due to collapse of Span. *alma* 'soul' and *arma* 'weapon'), Guaraní *pojoapy* contains *apy* 'deposit, extreme point,' Huambisa *manitai mani* 'warrior,' and Jarawara *tahi/tahi* also means 'killer,' 'hunter,' 'sliver,' and is also the name of a song about a particular spirit. Piro has a semianalyzable term with one constituent meaning 'man, person,' and Toba *n'ataxaqui* colexifies 'ammunition pocket' and 'spell.' Yanomámi *shëmotima* contains *shë* 'to hit,' and Hawaiian *mea make*, also meaning 'corpse,' is analyzable as 'thing die.'

97. The Window

Representation: 71%
Motivated: 44.6%
Thereof Analyzable: 36.9% Thereof Colexifying: 7.7%
Thereof by Contiguity: 19.5% Thereof by Similarity: 16.0%
Recurrent associated meanings: door, hole/opening, light/bright, see, house, mirror, glass, small, eye, wind/breeze, entrance

Terms for window are derived from verbs meaning 'to see,' as in Cubeo *jãĩcobe*, presumably analyzable as /jã́ino-cobe/ 'sight-CLASS.HOLE.LIKE.OBJECT,' in seven sampled languages, alongside Cubeo also in Muna, Sora, Upper Chehalis, Comanche, Kiowa, and Arabela, where the derivation mechanism is agentive nominalization and the resulting term hence can also refer to 'one who sees.' Furthermore, terms with 'see' of the lexical type are found in Khoekhoe (*mû-ǂui-dao-s* 'see-peep.through.opening-door-3SG.FEM'), Kiliwa (*?+wa?=t-p+sa-w-tay-u?* 'DN+house=SUBJ-MP+see-DIR-FREQ-OBL'), and Pawnee (*ukaʔaatawiiriku ~ ukaatawiiriku* /ukaʔaata-wi-iirik-hus/ 'shadow-??-see-IPFV'). The latter term colexifies 'window' with 'glass' in general and with 'mirror' in particular. Colexification with 'glass' is also found

(sometimes by analyzable terms) in Upper Chehalis, Oneida, Fijian (by the analyzable term *i ilo-ilo* 'DERIV look.at.reflection-RED), and Tuscarora; Nuuchahnulth has *pipiḥiqsim*, which is reduplicated from *piḥiqsim* 'glass, pane of glass,' Hawaiian has *puka-aniani* 'opening/door-glass/mirror/transparent,' and Yay *tu*[1] *cianʃ*[2] 'door mirror/glass.' Colexification with 'mirror' (again, sometimes by analyzable terms) is attested in Upper Chehalis, Kashaya, Kiowa, Pawnee, Tuscarora, Wintu (here the relevant term contains elements meaning 'in, be in, put in' inter alia and 'male person' which is related to a more general reconstructed root 'person'), and Fijian (by the analyzable term mentioned above). Associations with 'door,' as seen in the Yay term just mentioned, is, however, more frequent. It is realized by colexification in Itzaj, Cashinahua, Bororo, Jarawara, Hupda, Miskito, Fijian, Hawaiian, and Sedang, which colexifies also other meanings. Morphologically complex terms with 'door' are exclusively of the lexical type in the languages of the sample. There are a variety of subtypes. Commonly, 'small' or 'little' is the meaning of the second constituent in such terms, as in Toaripi *utape seika* 'door small,' and also in Biloxi, Kiliwa (where a third constituent 'house' is also featured), Abipón, Bororo, and Malagasy. Some of these terms can also be interpreted compositionally and hence also refer to a small door. Note also Chickasaw *okkisoshi'* /okkisa'-oshi'/ 'door-son' (and that the Ancient Greek word for 'window' is a diminutive of that for 'door,' Buck 1949: 470). Furthermore, Rotokas has *kiuvu ratao* /kiuvu rataoa/ 'wind door,' for which compare also Kanuri *kàsàm-rám* 'breeze-place.of,' Yoruba *ọjú-afẹ́fẹ́* 'eye-wind,' which precisely mirrors the etymology of English *windo*w, and Hawaiian *puka makani* 'opening wind' (this term also means 'opening for ventilation' and 'anus;' compare also the connetion of Spanish *ventana* with Latin *ventus* 'wind,' Buck 1949: 470). Khoekhoe has *mû-ǂui-dao-s* 'see-peep.through.opening-door-3SG.FEM,' Abzakh Adyghe *ṡḥe-n(ə)-ɣʷə-pśe* 'head/upper.part-RELAT/EPEN-??/end-door' (lexicographers note that the term is non-transparent nowadays), and Rotuman *nu'suar mutu* 'door cut.across/sever/cut.off.' In Greek, *paráthuro* is derived from *thúra* 'door,' and a semianalyzable term containing a morpheme meaning 'door' is featured in Yuki. Some of the terms colexifying 'window' with 'door' are morphologically complex, having constituents meaning 'house' and 'opening' or 'hole.' This is true of Itzaj (*jol-naj* 'hole-house'), Tuscarora (*yunęhsáhrarę* /yu-nęhs-ahr̲ar-ę/ '3SG.NEUT.PATIENT-house-be.a.hole-STAT'), Bororo (*wai-poro* 'house-opening'), and Hawaiian (*puka hale* 'opening house,' which also means 'door'). Kwoma has a term with an identical structure (*aka siisiiwey* 'house hole'), but uses it only for 'window,' not for 'door.' Associations between 'window' and 'hole, opening' are also found in other configurations, however (just as in Indo-European languages, Buck 1949: 470). Koyraboro Senni has *funtarey* 'pierce.hole-area.outside.' This term is at the same time a loanword from French *fenêtre* that is "vaguely intelligible" as a Koyraboro Senni compound according to the consulted source (that is, presumably, it has been folk-etymologized), Kolyma Yukaghir *pońqə-šeščə* 'light entrance' and *pońqə-söjnubəd-aŋil* 'light-enter-hole' (recall also that the means of derivation in the Cubeo term mentioned earlier is a classifier for hole-like objects, and note also that Hupda *məyó* is etymologically probably *məy-nə* 'house-mouth'), and Nuuchahnulth *kuukuuḥsim* contains *kuḥ*^{w} 'opening' and the lexical suffix *-sim* 'at an opening.' Furthermore, Kaingang has *kanẽ nor* 'eye opening,' Burarra, Cashinahua, Manange, and Sedang directly colexify 'hole, opening' with 'window,' and the Buli term also denotes

other kinds of openings in human-built structures. A semianalyzable term with this structure is furthermore found in Yay, and in Hani, *caqbyu* contains the classifier for holes, *byu*, with *caq* glossed as 'to guess, conjure up,' 'to have something break' inter alia. There are also terms in other languages which, as in Kolyma Yukaghir, betray an association with 'light' or 'brightness': Baruya *yi̇ri̇ti̇nna* /yi̇ri̇te-ti̇nna/ 'daylight-eye,' Chukchi *qeryɣəs-ʔən* 'light-one.who.has,' Cheyenne *vó'nâhenhestôtse*, containing *vó'n* 'light' and *hestôtse* 'thing,' Chickasaw *aahashtahli' ~ aahashtaali'* /aa-hashtahli-'/ 'LOC-be.bright-NMLZ,' Haida *radagats'aa7u* /radagaa-ts'a-7u/ 'be.daylight-inside-INSTR,' Lakhota *ožážąglepi*, the literal meaning of which is 'light frame' according to the consulted source, Central Yup'ik (Nunivak island dialect) *tanqiun* /tanqik-(u)n/ 'brightness-device.for' for a "seal-gut skylight window," and Miskito *ingni dimaika* 'light entrance' are of this kind (see Buck 1949: 470 for similar Germanic evidence; for the association with 'day,' note also that there is a semianalyzable term with a constituent meaning 'day' in Blackfoot). Moreover, Santiago Mexquititlan Otomí *neki xi hño* contains *neki* 'seem' and *hño* 'good' and also means 'clear, bright.' As seen in the Yoruba and Baruya terms already mentioned, metaphorical transfer from 'eye' to 'window' is another recurrent association. It is also attested in Mbum (*nják yâr* 'verge/entrance eye' and *yâr pàk* 'eye house') and Kaingang (*kanẽ nor* 'eye opening'), and there is a semianalyzable term in Jarawara (see Buck 1949: 470 for further evidence for this connection in Germanic, Slavic and Indo-Aryan). As has also become clear already from the discussion so far, 'house' is a frequent constituent in terms for 'window' as well. Alongside the terms already mentioned, terms containing an element meaning 'house' are also found in Carrier (*tadînṭaz*, containing *ta* 'house' and *ṭaz* 'gashing, slashing') and Bora (*jaatu páhoowa nééne* /jaá-tu páhoo-wa néé-ne/ 'house-ABL complete-CL.empty.space seem-CL.inan'),

Other associations include: Hausa *taga* also means 'to begin, attempt,' and *sagata*, which means 'crossbeam' inter alia, assumes the meaning 'window' in the dialects of Sokoto and Zaria. Dongolese Nubian *šibbā́g* also means 'lattice,' and the Mali term *manaingiēmgi* also means 'torch, flashlight.' Ngaanyatjarra *winta* is also used with the meaning 'glasses,' while Badaga *halagaṇṇi* also means "lookout, watchtower, observation point." In Ket *qoqpul* "small window made of ice in the winter dugout," the sequence *qo* may go back etymologically to *qō* 'ice.' Oneida *oˑwíseʔ* also means 'ice' (and 'glass'). Central Yup'ik *egaleq ~ legaleq* contains *ega-* 'to cook by boiling,' an association that is natural given that the term has been extended to 'window' from its original meaning "smokehole of a traditional Yup'ik house." Cayapa *juucapa* appears to contain *capa* 'point,' Fijian has the obsolescing term *i kasivi bale* 'DERIV to.spit drop' (with *i kasivi* meaning 'spittle'), and Samoan *(ʻo le) faʻa-malama* is analyzable as 'REL ART CAUS-blaze.up.' Finally, the Yay term *paak²taaŋ²* is analyzable as 'mouth different.'

98. The Adam's Apple

Representation: 47%
Motivated: 53.4%
Thereof Analyzable: 37.0% Thereof Colexifying: 20.0%
Thereof by Contiguity: 22.9% Thereof by Similarity: 28.5%

Recurrent associated meanings: throat/neck, larynx, windpipe, apple, bone, protrude/protuberance, fruit, trachea, goiter, knot

By contiguity, the most frequent association for the 'Adam's apple' is that with either 'throat' or '(front part of) neck.' Buli, Burarra, Dadibi, Kaluli, Muna, Wintu, Yuki, Fijian, and Hawaiian directly colexify 'Adam's Apple' with one of these meanings. There are many semianalyzable terms featuring one of the abovementioned meanings (namely in Bakueri, Rendille, Biloxi, Upper Chehalis, Comanche, and Macagúan). Fully analyzable complex terms with 'throat' include Khoekhoe *dom-ǃkhom-s* 'throat-bundle-3SG.FEM' (Khoekhoe also has *domo-ro-s* 'throat-DIM-3SG.FEM), Yoruba *kókó-ọ̀fun* 'wen/knob/excrescence-throat,' Comanche *pia kuitsi?* /pia kuitsi̱/ 'big throat,' Toaripi *kōvōre luka* 'neck.front/larynx stomach,' Haida *qagan skuji* 'throat bone,' Lesser Antillean Creole French *zo gòj* 'bone throat' (compare also Wappo *lèǩiš ċitiʻ-wélma*, which is perhaps analyzable as 'swallow bone-protection'), and Pawnee *raruucpaca*, analyzable as /raruus-paca/ 'throat bump,' while Guaraní colexifies 'throat' and 'Adam's Apple' by the analyzable term *jyryvikandu* /jyryvy-kandu/ 'throat-protuberance.' Similarly, Samoan has *pona-ua* 'knot/node/protuberance-neck,' Ineseño Chumash *šoqyokonič* is derived from *oqyokonič* 'be uneven, protruding' and Hawaiian colexifies 'Adam's Apple' with "[a]ny kind of protuberance from a pimple ... to a hill." The association with 'knot' is also found in another Austronesian language, namely Tetun, which has has *kaka-fukun* 'throat-knot/joint/knuckle.' Associations with 'apple' in this configuration are found in Basque (*zintzur-sagar* 'throat-apple') and Itzaj (*mansaanaj-il kal* 'apple-POSS neck/throat). Given that Itzaj *mansaanaj* < Span. *manzana*, as well as the fact that complex terms with 'apple' are found also in the other sampled European languages, namely Greek (*mḗl-on toú Adám* 'apple-NOM.SG.NEUT ART.DEF.GEN.SG.MASC Adam') and Welsh (*afal breuant* 'apple windpipe'), this pattern seems to be originally peculiar to Europe, see § 6.4.3.3. for historical evidence corroborating this conjecture.

Complex terms with 'neck' include Toaripi *kōvōre luka* 'neck.front/larynx stomach,' Yir Yoront *man-nhapn* 'neck-egg' and *man-pung* 'neck-sun,' Wintu *dokikenti,* perhaps analyzable as /dok-i-kenti/ 'neck-??-under,' Arabela *riquia-nu* 'neck-CLASS.PATH,' Cayapa *cutu pijpuca* 'neck wheel,' Chayahuita *cono-pira ~ conu-huira* 'neck-pile,' Guaraní *ajukytâ* /ajura-kytâ/ 'neck-wart' (alongside *ñoko'ê-kytâ* 'throat-wart'), Fijian *lobi ni domo* 'fold.lengthwise.or.crosswise POSS neck/voice/sound,' and Hawaiian *kani-ʻāʻī* 'hard-neck' (which however, can refer to the 'neck' itself inter alia; another Hawaiian term, *pūʻāʻī*, probably goes back to **puʻu ʻāʻī* 'protuberance neck'). Semianalyzable terms are found in One, Yir Yoront, Nivkh, Yana, Toba, and Yay.

Lavukaleve colexifies 'Adam's apple' with 'fruit,' Sedang has *plai plê* 'fruit/seed/pellet armspan,' and a semianalyzable term with 'fruit' is attested in Lenakel. Khoekhoe, Dongolese Nubian, Burarra, Muna, Badaga, Khalkha, Aguaruna, Arabela, Carib, Cashinahua, Fijian, and Hani colexify 'Adam's Apple' with 'larynx' (Upper Chehalis has a semianalyzable term with this being the meaning of the identifiable constituent), Kyaka, Wintu, and Carib with 'windpipe,' and Chayahuita with 'esophagus.' Miskito *won krukmaya* consists of *krukmaya* 'goiter' and the classifier *won*, and Wayampi colexifies the relevant

meanings directly. Kyaka colexifies 'Adam's apple' with 'trachea,' and Wintu *holhol* is reduplicated from *hol* 'pipe, tube, trachea' (also meaning 'light, bright, shine').

Other associations include: Dongolese Nubian *gőz* also means 'dune,' though in this meaning it is borrowed from Sudanian Arabic. Dadibi *dogoni* also denotes the 'sound of voice,' and the Kwoma term *noku pipoy* appears to be analyzable as 'sago.palm strike.down.' Abzakh Adyghe *qʷʔəmbəqʔay*, colexifying 'glottis,' contains either *qʷʔə* 'skeleton' or *(-)qʷʔ(e)* 'pointed object,' the epenthetic element *-m-*, and *b(e)*, which occurs in a number of complex terms referring to the human body. Badaga *gaṭṭigallu* is analyzable as /gaṭṭi-kallu/ 'cartilage-stone,' and Cheyenne *-nêhpo'hoo'ôtse* as /-nêhpo'ohe-hestôtse/ 'locked-thing' (the literal translation given in the source is 'valve/thing that closes'). The Chickasaw term *inonkopoolo'* is perhaps analyzable as /inonka' oppolo-'/ 'voice be.ruined-NMLZ.' Piro *swayipla* appears to contain *swa* 'orifice,' Toba colexifies 'Adam's apple' with 'gill,' and the Yanomámi term *krukupɨ* is also used to refer to the 'hyoid bone of the Araguato monkey.' Fijian *i tagi-tagi* is analyzable as 'DERIV make.sound-RED,' Hani *kaoqciivq* might be related to *ciivq*, meaning 'to choke, pinch' and 'genealogy,' alongside acting as a classifier for 'joints,' for instance on the fingers or on bamboo. Samoan colexifies 'Adam's apple' with "eyes of a snake or eel" and 'uvula.'

99. The Ankle

Representation: 86%
Motivated: 46.1%
Thereof Analyzable: 41.7% Thereof Colexifying: 6.4%
Thereof by Contiguity: 13.5% Thereof by Similarity: 29.2%
Recurrent associated meanings: foot/leg, joint/wrist, eye, neck/throat, knot, bone, egg, heel, (Achille's) tendon, stone/pebble, seed

Motivated terms for 'ankle' are overwhelmingly morphologically complex, with 'foot' and/or 'leg' being the meaning of one of the elements. Mostly, the semantics of the second element is such that it denotes a small roundish object, thus constituting a metaphorical transfer. However, this is not always so clear, and in fact the most frequent subtypes attested in the sampled languages do not adhere to this pattern.

In thirteen sampled languages, Swahili, One, Toaripi, Badaga, Lake Miwok, Kaingang, Miskito, Wayampi, Hawaiian, Kapingamarangi, Mandarin, White Hmong, and Bislama, 'joint' or 'wrist' is the meaning of the second constituent (the Hawaiian term colexifies 'heel,' as does a simplex Badaga term). For one, One has *teu tampla* 'foot/leg joint;' Welsh, Cahuilla, Haida, Wintu, Rama, and Tehuelche colexify 'wrist' or 'joint' with 'ankle,' Rama also colexifies 'knuckle' specfically. The next most frequent subtype is that where the meaning of the second constituent is 'neck' and/or 'throat,' as in Sahu *rou ma camala* 'foot/leg POSS neck,' clear cases of which are found in nine sampled languages, alongside Sahu in Efik, Yoruba, Sko, Abzakh Adyghe (where 'neck' is colexified with 'top'), Japanese, Itzaj, White Hmong, Vietnamese, and possibly Bakueri, though this is unclear (the Sko term stands out in that here not 'neck,' but 'neck bones' specifically is the meaning of the constituent, perhaps providing a clue to the underlying perceived similarity responsible

for this strategy in general). There are also associations with 'bone' in other languages: Santiago Mexquititlan Otomí has *bots'undo'yo wa*, containing *ndo'yo* 'bone' and *wa* 'foot,' Rotokas and Cubeo have terms with a similar structure and Cubeo also others, in which the motivation is constituted by classifiers; Copainalá Zoque has a semianalyzable term where the identifiable constituent is bone.'

A more clearly metaphorically motivated pattern of transfer from one ("salient") body-part to another (less "salient") one is constituted by complex terms where the meaning of the second constituent is 'eye,' as in Koyraboro Senni *čee-moñe* 'foot-eye.' Such terms are also found in Hausa, Kanuri, Anggor, Kwoma, Kyaka, Kosarek Yale, Pipil, and Hani. Moreover, Piro *yhale-xi* is analyzable as 'eye-fruit/seed/DIM' (the term colexifies, in addition, 'leaf bud' and 'belt buckle'), and Sedang directly colexifies 'ankle' with 'eye' inter alia. Again, one term stands out, namely that in Kosarek Yale, where an additional constituent meaning 'fruit,' 'seed,' and 'egg' inter alia (see § 6.4.3.13.3. for this pattern of colexification) is present. A semianalyzable term with 'fruit,' colexifying 'kidney,' 'side of ankle' and 'earlobe,' is also found in Lenakel, and an association between 'ankle' and 'egg' is also revealed by the Chayahuita term *ca'yo-pira* 'egg-pile' and is also present in Bwe Karen, which has *-khaʃí dî*, containing *-kha* 'leg, foot' and *dî* 'egg, testicle.' Common are also complex terms with 'foot' and/or 'leg' being the meaning of one constituent, and 'knot' that of the other, as in Badaga *gaṇṇu ka:lu* 'knot/joints leg.' Such terms (with or without colexification with 'joint') are also attested for Pawnee, Tsafiki, and Tetun, direct colexification is attested in Tehuelche, and a semianalyzable term with one constituent meaning 'knot' or 'joint' is found in Rotuman. Further, Dongolese Nubian has *óssiŋkugúndi ~ óssiŋkumúndi* /óssi-ŋ-kugúnd(I)/ 'foot-GEN-elbow,' Aymara has *kayu moqo* 'foot pebble,' and Arabela colexifies 'ankle' and 'stone, rock.' Further complex terms with one of the constituents being 'foot' or 'leg,' many of them reconcilable with the general tendency of metaphorical transfer of small roundish objects, are Yoruba *kókosè*, analyzable either as /kókò-ẹsè/ 'edible.tuber-leg' or /kókó-ẹsè/ 'knot.in.tree/knob/excrescence-leg,' Sko *làngòe*, presumably analyzable as /làng-óe/ 'foot-yam,' Abzakh Adyghe *λeč̣ʔen* /λ(e)-č̣ʔen/ 'foot-ossicle/lump' (a semianalyzable term containing an element which can refer to a 'lump' inter alia is found in Rotuman), Biloxi *sponí ~ i´sponí ~ iñksponí ~ sponitú ~ i´sponitú ~ iñksponitú* /asi-po-ni´/ 'feet-wrap.up.in.bundle-CAUS,' Chickasaw *iyyi̲mosak* /iyyi' imosak/ 'foot DAT-hickory.nut' (the analysis is marked as questionable in the source), Highland Chontal *galgoxac gahmis* /galgoxac lahmis/ 'shelled.corn leg,' Kashaya *qʰama pʰiǩo* 'foot ball,' Imbabura Quechua *chaki muku* 'foot nude,' Wayampi *pɨyũʔã* /pɨ-yu-ʔã/ 'foot-spine-maintain.upright.position' (this term colexifies 'claw of bird of prey'), Bwe Karen *-kha-dé* '-foot-narrowest.part,' Samoan *tapu-vae* 'forbidden-leg/foot' (colexifying 'Pig's trotters'), Takia *ŋe-n ŋdu-n* 'leg-3SG nose-3SG,' Tetun *ain-liras* 'foot/leg-wing/fin,' and Yay *ho⁴θay¹ tin¹* 'adam's.apple foot.' Semianalyzable terms of this kind are found in Mbum, Kaluli, Kemtuik, Ngaanyatjarra, Rotokas, Sentani, Yir Yoront, Badaga, Carrier, Kiliwa, Kiowa, Oneida, San Lucas Quiaviní Zapotec, Guaraní, and Hani. Aguaruna colexifies 'ankle' and 'tendon,' and Yir Yoront colexifies 'Achille's tendon' specifically (Guaraní, in addition, has a semianalyzable term for the 'ankle' where the meaning of the identifiable constiutent is 'Achille's tendon').

Other associations include: Khoekhoe *!goros* also denotes the 'fetlock joint of animals,' and Rendille, presumably due to homonymy, colexifies the "self, human ego." Ngaanyatjarra *warungantjil(pa)* also denotes a "solid log that continues burning a long time" and the 'Australian Hobby' (a species of falcon). Rotokas *arioisi* also means 'kapiok seed.' Khalkha *šaγa(n) ~ siγa(n) ~ siγai* may be related to *šaγa- ~ siγa-* 'to hit with fist, drive a nail or peg in.' Welsh *swrn* also means 'fetlock' and 'a good number.' Ineseño Chumash *šow* colexifies 'ankle' with 'tobacco,' while Lesser Antillean Creole French *chivi* also means 'plug, peg' and 'pin.' Central Yup'ik *cingilleq* is derived from *cingiq* 'shoelace' by means of the postbase (see § 4.4.2) *-lleq* 'former.' Bora *lleébou* might be related to a verb meaning 'to listen, obey,' Cayapa *nemiipijpuca* contains *puca* 'small round thing,' while Macaguán *peatá* appears to be derived from *atá* 'hard' by prefixation of a possessive marker. The Jarawara term *rabi* also means 'pencil' due to Portuguese influence, while Ancash Quechua *utsu putu* is analyzable as 'garlic container' and indeed also denotes a vessel in which garlic is stored and which is similar in appearance to the ankle. Toba *nqonacamo* appears to contain *qona* 'toe, claw,' and Malagasy *kìtro kèly* is analyzable as 'hoof small.'

100. The Beard

Representation: 92%
Motivated: 51.8%
Thereof Analyzable: 33.5% Thereof Colexifying: 19.9%
Thereof by Contiguity: 27.5% Thereof by Similarity: 15.9%
Recurrent associated meanings: hair, chin, mouth, whiskers, fur, feather, jaw, lip, antenna, beard of grain, face, corn silk, barbel, wool

The most frequently associated meaning for 'beard' (without differentiating between 'beard,' 'mustache,' 'goatee,' etc.) is, unsurprisingly, 'hair.' Kosarek Yale, Tasmanian (Northeastern), Nivkh, Bororo, and Ancash Quechua colexify 'hair' and 'beard' directly; otherwise, the association is realized by complex terms, the other element having a variety of meanings in the languages of the world. Most frequently, in Mali, Toaripi (for 'mustache' specifically), Sora, Cheyenne, Kiliwa, Nuuchahnulth (with the element referring to 'mouth' being a lexical affix), Pawnee (where 'be hairy' is colexified with 'be furry, fuzzy'), Pipil, Wintu, Copainalá Zoque, Lenakel, Malagasy, and Tetun, the second constituent makes reference to the 'mouth,' as in Tetun *nunun-rahun* 'mouth/lip-body.hair,' which also denotes a 'goatee' specifically (the constituent means 'hairy' rather than 'hair' in the case of Cheyenne; this situation is recoverable etymologically in Cahuilla, compare also Hani *meiqmoq* with *meiqbaoq* 'mouth, snout' and *aqmoq* 'horse, pubic hair'). There are, however, also complex terms of the lexical type with 'mouth' where the second element has a meaning other than 'hair:' Khoekhoe has *lhō-ams* 'wrinkle-mouth,' Ngambay *mbáy-tà* 'lord-mouth/lips,' Abzakh Adyghe *żač'e* /że-č'e/ 'mouth-end,' Cashinahua *kex-ni* 'mouth/lip/edge-forest,' Hupda *nɔ-cúg* 'mouth-hummingbird,' and Takia *awa-n dabi-n* 'mouth-3SG root-3SG' (it is possible that clearly metaphor-based terms as in the latter two languages are due to dense beards being an item of "acculturation" so to speak - growth of bodily hair is less pronounced for instance among the indigenous population of the Amer-

icas when compared with Europeans). In Buin and Yir Yoront, there are semianalyzable terms for 'beard' featuring a constituent with the meaning 'mouth.'

Also very common are terms of the analyzable type with the constituents meaning 'hair' and 'chin,' as in Berik *olele safa* 'chin hair/fur/feather.' Such terms are also attested in Efik, Dadibi, Kwoma, Meyah, Toaripi, Lakhota, Rama, Toba, Wayampi, Great Andamanese, and Bwe Karen. Other complex terms on the basis of 'chin' are Katcha *musá mo buruŋɛ ~ musá mo böröŋe* 'wool GEN chin' (similarly, the association with 'wool' is present due to colexification with 'hair' in the Toba term mentioned above), Waris *keu-ta* 'chin-small.object,' and Bora perhaps *újca-he* 'chin/beard-CL.oblong,' but segmentation is unsure. 'Chin' and 'beard' are directly colexified in Dongolese Nubian (also with 'whiskers,' a pattern of colexification also attested in Baruya, Nunggubuyu, Rotokas, Western Tasmanian, Carrier, Chickasaw, Ineseño Chumash, Haida, Kiowa, Lake Miwok, Pawnee, Xicotepec de Juárez Totonac, Wappo, Wintu, Central Yup'ik, San Lucas Quiaviní Zapotec, Bislama, where it is archaic, and Hani, where 'chin whiskers' more specifically is colexified), Gurindji, Ngaanyatjarra, Northeastern, Middle Eastern, and Western Tasmanian, Badaga, Embera, Fijian, Rotuman, and Tetun; a semianalyzable term where the identifiable constituent is 'chin' is featured in Chukchi and Middle Eastern Tasmanian (the association is common in Indo-European, Buck 1949: 205). Kwoma and Rama colexify 'chin' with 'jaw,' and therefore their terms for 'beard' also feature an association with 'jaw' more specifically. Complex terms with 'jaw' alone and 'hair,' such as Yaqui *chao boa-m* 'jaw body.hair-PL,' are also found in Yoruba, Sko, and Chickasaw. Miskito colexifies 'beard' with 'jaw' (perhaps a compound with 'jaw' is present in Ancient Greek, Buck 1949: 205). Baruya has *maryaata* /marya-mɨjata/ 'lip-hair,' and terms with such structure are also found in Ngambay, Abipón, Piro (where the constituent making reference to the lips means 'rounded lips' more precisely), and Tetun, as already noted above, while Cashinahua has, as also already noted above, *kex-ni* 'mouth/lip/edge-forest.' Jarawara has *noki baki kone/noko bako kone* 'face hair' for 'beard on side of face,' and terms with such structure (without a semantic restriction mentioned like that in Jarawara) occur also in Nuuchahnulth (with the 'face' component expressed by a lexical affix) and Wintu (where 'face' and 'eye' are colexified). Other complex terms where one of the constituents is 'hair' are Kiowa *sęįn-p'ą̄'-gɑ* 'mucus.of.nose-body.hair-NOUN.POSTFIX' and Hawaiian *hulu weuweu* 'body.hair grass/bushy' for "[d]owny feathers or beard, fuzz." A semianalyzable term is found in Japanese.

Now, the above discussion does not take into account the lexical differentiation made in many languages between 'hair of the head,' 'body hair,' and potentially even more fine-grained distinctions. Since 'body hair' is frequently colexified with 'fur of animal' and/or 'feathers,' a number of languages betray associations also with these meanings. With 'fur,' this is the case by colexification in Kosarek Yale and Ancash Quechua and by analyzable terms in Berik, Kwoma, Meyah, Toaripi, Chickasaw, Kiliwa, Lakhota, Nuuchahnulth, Pawnee, Wintu, Toba, Bwe Karen, Lenakel, and Tetun. In two sampled languages, Kolyma Yukaghir and Kashaya, it is specifically 'fur' rather than 'hair' which is the constituent of complex terms. Note also Tehuelche *ʔašč'ex ~ ač'ex ~ ʔačx ~ ʔač'ex ~ ʔaščex ~ a:č'ex ~ a:čex* 'in-hair/fur,' which colexifies 'beard' with 'eyebrow' and 'eyelash.' 'Beard' is associated by analyzable terms with 'feather' in Efik, Berik, Kwoma, Meyah, Toaripi,

Chickasaw, Toba, Bwe Karen, Hawaiian, Lenakel, and by colexification in Kosarek Yale and Bororo (sometimes also with further meanings, cf. section 21).

Furthermore, Xicotepec de Juarez Totonac, Arabela, Cavineña, and Piro colexify 'beard' with 'antenna of insect,' Basque, Khalkha, and Nez Perce with 'beard of grain,' Basque and Hawaiian with 'barbel,' and Kiowa and Piro with 'corn silk' specifically.

Other associations include: Hausa *gemu* also denotes the concepts 'corbel' and 'superior person,' while Koyraboro Senni *kaabe ~ kaaba* is also the name of a spice and a tree species. Gurindji *jika* is also the term for the 'brush-tailed rat kangoroo' and ornaments made from its tail. Kwoma colexifies beard with "shell and feather decoration glued on the jaw of certain ceremonial sculptures carved in the form of heads," and Muna has *samba ~ kasamba* for "beard (on chin and cheeks)," which is borrowed from Indonesian; indigenously, *samba* also denotes a kind of girdle inter alia. Khalkha *ziber* also means 'fins of fish' (the meaning 'beard' is register-specific), Nivkh colexifies 'beard' with 'vegetation,' 'feeler,' and 'tentacle' (the term is very short, so homonymy is a possibility). Cahuilla *támaš* is also used to refer to a bearded person. Santiago Mexquititlan Otomí colexifies 'beard' and 'flour,' while Guaraní *tendyva* contains *tendy* 'saliva.' Ancash Quechua *sapra* is also metonymically capable of referring to a 'Spaniard.' Fijian *kumi* also denotes 'Tongan cloth,' Kapingamarangi *tuu i lala* seems to be analyzable as 'cut at down,' while Rotuman *kumkumu* also denotes a species of crab. Bislama *mustas* (< Engl. *mustache*) is also the name for the 'goatfish.'

101. The Bladder

Representation: 59%
Motivated: 67.2%
Thereof Analyzable: 47.3% Thereof Colexifying: 19.9%
Thereof by Contiguity: 31.6% Thereof by Similarity: 22.4%
Recurrent associated meanings: urine/urinate, container, bag, balloon, place, blister, gall bladder, bubble, belly/stomach, sack, house, goiter, eye, inflate

Unsurprisingly, terms for the 'bladder' frequently contain an element meaning either 'urine' or 'to urinate.' 'Urine' / 'urinate' and 'bladder' are colexified directly in Anggor, Ngaanyatjarra, Tuscarora, Aguaruna, Huambisa, Yanomámi, and Bislama. Derived terms, such as Ineseño Chumash *šoxšol,* derived from *oxšol* 'to urinate,' are also found in Bakueri, Hausa, Upper Chehalis, Abipón, Bora, Chayahuita, Cubeo, and Toba. In terms of the lexical type, 'container' is most frequently the meaning of the second constituent, as in Ket *dɤsol* /dɤ̄s-óòl/ 'urine-container,' and also in Chickasaw, Kashaya, Kiliwa, Nuuchahnulth, Bororo, Carib, Kaingang, and Ancash Quechua. Similarly, the Toba term is derived by a classifier for receptables. A specific container, namely 'sack,' is also attested as the meaning of the second constituent, as in Mbum *fã̄-jáù* 'sack-urine.' Such a term is also found in Ket, and 'sack' and 'bladder' are colexified directly in Nez Perce inter alia. Very similarly, Yoruba has *àpo-ìtọ̀* 'bag-urine,' and so do Carrier and Kiowa, while the association is realized by colexification in Piro and Hawaiian (furthermore, Tuscarora has an optional complex term of this kind on the basis of its term colexifying 'bladder' and 'urine').

Semianalyzable terms where the identifiable consituent means ‘sack’ or ‘bag’ are found in Highland Chontal and Samoan. Terms with ‘house’ (which is a kind of container of sorts as well), such as Embera *síua-dé* ‘urine-house,’ are also attested in Mbum and Miskito (in the latter language, there are also alternative terms where ‘house’ is replaced by ‘womb’ and ‘urine’ by ‘liquid’ as a constituent). More generally, ‘place’ is the meaning of the second constituent in Arabela *shaaca quiquio* ‘urine place,’ as well as in Lake Miwok, Miskito, and Great Andamanese. Rama *ngústi úp* is analyzable as /ngústi up/ ‘pissing eye,’ while Kyaka has *puu renge* ‘urine eye’ (both *puu* and *renge* have other meanings). Other complex terms where ‘urine’ is a constituent are Buli *sinsam-liuuk* ‘urine-passage/hole/path,’ Efik *u-tök ikïm* ‘NMLZ-to.void urine,’ Katcha *nimo misərə* /nimo m-kisərə/ ‘thing ??-urine,’ Cheyenne *xaenoo'ôtse*, analyzable as /xae-noo'ohtsé/ ‘urinate-depart,’ Oneida *-nhʌhalákhwaʔ*, containing the roots *-nhʌh-* ‘urine’ and *-l-* ‘be in or on,’ Pipil (Cuisnahuat dialect) *-a:xi:x-tekun* ‘-urine-bottle.gourd,’ Yuki *as´-pis´* ‘urine-from,’ San Lucas Quiaviní Zapotec *gui'dy x:qui'ix* /gui'ihdy x:qui'ix/ ‘skin urine,’ Tehuelche *t'ep'enʔašk'en*, containing *t'ep ~ t'e:p ~ t'ep' ~ t'epe ~ ʔet'ep* ‘urine, urinate’ and *-ašk'en* ‘interior,’ and Hawaiian *puʻu-mimi* ‘protuberance-urine.’ Semianalyzable terms are found in Bakueri, Kaluli, Chukchi, Biloxi, and Guaraní. Wintu *pʰuˑλemes* and *pʰuˑλes* contain *pʰuλ* ‘to blow,’ the former colexifying ‘bologna, meat in casings, sausages’ (note also Fijian *i uvu-uvu* ‘DERIV blow.with.mouth-RED’).

Basque, Embera, and Macaguán colexify ‘bladder’ with ‘blister,’ Ngambay and Hawaiian with ‘goiter’ (Ngambay also with ‘evacuate’ and ‘gall bladder,’ it shares the latter association with Ngaanyatjarra and Yir Yoront, and Hawaiian also colexifies ‘womb’ among other meanings), while Pawnee and Cavineña have terms derived from verbs meaning ‘to inflate’ (*paaruris*, analyzable as /waarur-his/ ‘inflate-PERF’ and *cujacuja*, reduplicated from *cuja* ‘to inflate’ respectively; both terms colexify, as do Lesser Antillean Creole French and Samoan, ‘balloon’). ‘Bladder’ and ‘stomach’ are colexified in Nuuchahnulth and Hawaiian, and similarly, Lenakel has *nouanetpɨnami-* /noua-netp-nami-/ ‘fruit-belly-urine-,’ which colexifies also ‘ball.’ Finally, Itzaj, Cayapa, and Hani colexify ‘bladder’ with ‘bubble’ (Cayapa colexifies ‘animal bladder’ more specifically, also with ‘foam’).

Other associations include: Buin *kou* is a general term for the internal organs such as ‘guts,’ including the bladder. Kwoma *mokugwey* appears to contain *moku* ‘semen,’ and Chukchi *pəɣəlqewəcʔen* contains *pəɣəl-* ‘float.’ Greek *kýstis* also means ‘cyst,’ and Khalkha *dabusang ~ dabasaɣ ~ dabusaɣ ~ dabisaɣ* is also used to refer to the ‘lower part of the abdomen’ and ‘the pubic region.’ San Mateo del Mar Huave *mipeparan* seems to contain the verb *apep* ‘to catch.’ Itzaj *b'ejiigaj* also means ‘innertube,’ while Nez Perce colexifies ‘bladder’ with ‘shell’ and ‘coffin.’ Xicotepec de Juárez Totonac *a'nlhā tāquīˈ xcān* appears to be analyzable as ‘where get.up water,’ and Central Yup’ik *nakacuk* might contain the postbase (cf. § 4.4.2.) *-cuk* ‘unpleasing, bad one.’ The Aymara term *llaq'allachi ~ yaq'allachi* may contain *yaq'a* ‘excrement,’ while Guaraní *tyryru* also means ‘urinal.’ Hawaiian *koʻana* also is used with the meaning ‘dregs, sediment.’

102. The Blood

Representation: 98%
Motivated: 15.4%
Thereof Analyzable: 2.4% Thereof Colexifying: 13.0%
Thereof by Contiguity: 7.4% Thereof by Similarity: 1.5%
Recurrent associated meanings: red, resin, race/kinship, menses/menstruate, foetus, body liquid, liquid

Terms for 'blood' are not frequently motivated. The most frequent association (ignoring ambiguous terms with respect to word class which can hence also mean 'to bleed'), occurring in nine sampled languages, is that with the color 'red,' by colexification in One, Toaripi ("blood red" more specifically), Wintu (colexifying also "arse"), Aymara, Rama, and Hawaiian, and by complex terms in Abipón and Bora (*n-aoig-Ra* 'POSS.INDEF/3SG-red/yellow-CAUS' and *túpja-cyo* /túpjá-cyo/ 'red-CL.pointed' respectively). Yanomámi *iyëiyë ~ ɨyëɨyë ~ ĩyëĩyë* is reduplicated from *iyë*, which means 'bloody,' 'blood-colored,' and 'red.' The association with 'red' is also found in Sanskrit (Buck 1949: 206). A semianalyzable term is furthermore found in Yana, and the association is etymologically recoverable also in Dongolese Nubian. Colexification with 'resin, sap' is found in Kwoma, Cubeo, Maxakalí, Fijian, and Hawaiian (note also Sko *hì* 'blood' and *hí* 'sap' and the colexification of 'blood' with "red sap of certain trees" specifically in Jarawara). 'Blood' and 'body liquid' generally are colexified in Kyaka and Hawaiian (in Kyaka, the term also denotes 'lymph' and 'serum,' and, presumably by homonymy also means "song, singsang, chant"), and, still more generally, 'blood' and 'liquid' are in Bororo (among other meanings), Maxakalí, and Hawaiian. In Wayampi, *tɨɛʔɛ* contains *tɨɨ* 'juice, liquid,' and the other element is said to etymologically have the meaning 'true, genuine.' Another Wayampi term, *tuwɨ*, also means 'dry.' Furthermore, 'blood' is colexified with 'menses' or 'to menstruate' in Dongolese Nubian, Lake Miwok, and Samoan (where the relevant term is reduplicated from *'ele* 'red soil' and also denotes the 'earth' and the 'dirt'), with 'race' or 'kinship' in Basque, Lesser Antillean Creole French (inter alia), Itzaj, Ancash Quechua, and Rotuman, and with 'foetus' in Efik and Tetun.

Other unique associations in the languages of the sample include: Efik *iyip* is also used with the meaning 'murder, bloodshed,' and Hausa *jini* is also the name of a children's game. Kwoma is unique in using the same term, *pi*, for 'blood' and 'milk' inter alia, and Muna *rea* also means to "have a communal garden." Waris *tóvól ~ nihtóv* also means 'skin,' and Kosarek Yale *eneng* also 'pus.' Abzakh Adyghe *λə* also means 'be lying, suspended, situated' inter alia, and Haida *ray* is also used with the meaning 'kidney of salmonid.'' For *utkwà·reh ~ ú·tkwareh*, a Tuscarora term for 'blood,' the lexicographer remarks that the underlying root *-tkwar- ~ -tkwa̲r-* might be an old compound of *-tkw-* 'stomach' and *-r-* 'to be in.' Central Yup'ik *qayuq* also means 'soup, broth,' and Copainalá Zoque *nʌpin* probably contains *nʌ'* 'water.' Toba *ltago'q* also means 'one's own son.' Bwe Karen colexifies 'to steam' inter alia and by another term also 'nest,' and Hawaiian *koko* also means 'rainbow-hued' inter alia.

103. The Bone

Representation: 99%
Motivated: 21.7%
Thereof Analyzable: 0.7% Thereof Colexifying: 20.3%
Thereof by Contiguity: 6.1% Thereof by Similarity: 11.0%
Recurrent associated meanings: skeleton, stalk/fibre, seed, strength, awl/needle, hard/hardness, leg, tree, horn, fishbone, shell, dry, corpse

Like 'blood,' 'bone' is a meaning for which few lexical associations are found cross-linguistically. The most frequent one, by configurational contiguity, is that with 'skeleton,' occurring by colexification in Khalkha, Tuscarora, Bororo, Cashinahua, Embera, Guaraní, Huambisa, Tehuelche, and Rotuman. Further, Abzakh Adyghe has *qʷʼə-mə-śḥe* 'skeleton-RELAT/EPEN-upper.part/??.' Khalkha inter alia colexifies 'bone' with 'corpse,' and Hawaiian with 'carcass,' as of chicken. In Wappo and Bororo, relevant terms also mean 'hard' or 'hardness' (note also Rotokas *kerua*, presumably containing *keru* 'harden like bone' and a classifier for narrow objects and Sentani *bo* 'bone' and *bobo* 'hard'). Cheyenne and Lengua colexify 'bone' with 'leg' (note also the cognacy of German *bein* 'leg' and English *bone* as well as a further association of this kind in Greek noted by Buck 1949: 207; perhaps this pattern is motivated by the fact that the thigh bone is the largest bone in the human body?). Wappo and Yuki, presumably by functional contiguity, use the same word for 'bone' and 'awl' or 'needle.'

There are also a number of metaphor-driven associations, although these are not very frequent as well. Baruya, Ngaanyatjarra, Toaripi, Sahu, Sora, Arabela, Lengua, Maxakalí, and Yanomámi colexify 'bone' with 'stalk' or 'fibre' (Yanomámi also with 'cotton' and 'animal fur'). This pattern is also found in Indo-European as evidenced by cognates in Baltic, Greek, and Latin (Buck 1949: 207). Maxakalí and Tsafiki colexify 'bone' with 'tree' (as well as 'stick' or 'pole' in Tsafiki), Yir Yoront, Basque, Khalkha, and Wintu (among other meanings in the latter two languages) with 'seed' (this pattern is also attested in Indo-European, Buck 1949: 207), Highland Chontal and Itzaj with 'horn' (the latter language also with 'employment'), and Buli and Basque with 'fishbone.' San Lucas Quiaviní Zapotec *zu'aht* also denotes the 'shell,' e.g. of a turtle (alongside a rattle used in a particular dance), and Hawaiian *iwi* also colexifies 'bone' with 'shell,' among other meanings. In Ngambay, Baruya, Kwoma, and Waris, relevant terms also mean 'strong' or 'strength,' see Aikhenvald (2008: 579) for brief discussion of this association in Manambu suggesting areal convergence. In Anggor, *hamɨndɨ* is also glossed as 'very,' which suggests that it is used as an intensifier. Finally Guaraní *kâ* also means 'dry' (and 'pit of fruit'), and Burarra (-)*jorla,* meaning 'bone' in the Gun-nartpa dialect, generally also means 'dry' (as well as 'sun-baked' and 'no water').

Other associations include: Efik *ök'pö* also denotes a kind of yam, the fruit of a particular tree, a creeping vine, and 'bird lime.' The Burarra term *-mama* also denotes the 'frame of a structure' and a 'coin' as opposed to paper money. Dadibi *dili* also means 'singsing.' Gurindji *kuyuwarn* contains *kuyu* 'meat,' whereas the Kwoma term *hapa* also means 'bone of the upper arm' specifically and then by extension

also ‘upper arm.’ The Ngaanyatjarra term *tarrka* also means ‘bare area’ as well as ‘skinny’ in an adjectival sense. Greek *kókkalo* also may refer to ‘ivory on the piano key’ and ‘shoehorn,’ and Ket *aˀt* also means ‘soon.’ As a verb, Kiowa *t‘ęį’m* means ‘to break.’ Lake Miwok *kúlum* also means ‘cemetery,’ and Macaguán *-chit* also ‘shinbone.’ Hani *saqyyuq* and *saqyoq* may be related to *saq*, meaning ‘muscle, meat’ inter alia, Bwe Karen *khwi* is also the name of a sand lizard, the Kapingamarangi term *iwi* also means ‘upper ridge pole’ as well as ‘crease’ and ‘sharp edge,’ and the cognate Samoan *iwi* also means ‘limb.’ Rotuman colexifies ‘bone’ with “projecting spike … on the fins of certain fish” inter alia, Takia *tatu* also means ‘spine,’ while Tetun *ruin* also means ‘rough,’ and Sedang *kɤxiang* may or may not contain *kɤxi* ‘egg.’ Yay *dok²* colexifies ‘flower’ and *ku⁵* also means “pair, even (of numbers).”

104. The Brain

Representation: 90%
Motivated: 33.8%
Thereof Analyzable: 20.2% Thereof Colexifying: 13.7%
Thereof by Contiguity: 9.0% Thereof by Similarity: 22.2%
Recurrent associated meanings: head, marrow, mind, pus, semen, pith, faeces

Frequently, motivated terms for the ‘brain’ are morphologically complex and contain a constituent meaning ‘head’ (see Buck 1949: 215 for similar evidence from Indo-European). As for the semantics of the second constituent, ‘marrow’ is well attested. For instance, Kaingang has *krĩ kujo* ‘head marrow,’ as do Koyraboro Senni, Berik, and Kiliwa. ‘Brain’ and ‘marrow’ are, however, also frequently directly colexified. This is the case in Yoruba, Burarra, Kyaka, Yir Yoront, Chickasaw, Central Yup’ik, Arabela, Bororo, Ancash Quechua, Toba, Yanomámi (where *heoshipë ~ heyõshipë* appears to contain *heõshi* ‘hot’ and the quantal classifier *pë*, for which see § 4.4.1.1.), Hawaiian (alongside other meanings), Lenakel, White Hmong, and Yay, while Bwe Karen *í*, colexifying ‘to give,’ is said to appear to be an alternant of *i* “soft centre (of a plant or tree), marrow (of a bone), etc.” Furthermore, Great Andamanese has a derived term from a root colexifying ‘marrow’ and ‘pus’ (the association with ‘marrow’ is not at all unknown, see e.g. Buck 1949: 215 and Jóhannesson 1949: 88 for for Indo-European, Matisoff 1978: 233 for Southeast Asian, and Wilkins 1996 for Australian evidence; note in this context also the quote from Aristotle mentioned by Buck 1949: 215: “for many think the brain is really marrow”). Great Andamanese, in fact, colexifies ‘pus’ and ‘marrow,’ and an association with ‘pus’ is also found for ‘brain’ by colexification in Ket and Lake Miwok and by the analyzable terms *wi-dɔ̀* ‘pus-head’ in Ngambay and *lal mabiara* ‘head pus’ in Miskito (note also Kiliwa *?iy=xtla?=pii* ‘head=??=puss,’ with ‘puss’ a mistake for ‘pus’?). Basque has *burmuin*, presumably analyzable as /buru-muin/ ‘head-pith/bud/shoot/sprout’ (note that there is a complex term on the basis of*muin* meaning ‘marrow’), and ‘brain’ and “pithy core of any tree” are colexified in Nuuchahnulth (alongside ‘heart’ and ‘spinal cord’). Other analyzable terms with ‘head’ being one constituent are Buli *zupuuk* /zuk-puuk/ ‘head-foam,’ One *sila silla* ‘head leaf,’ Abzakh Adyghe *śḥekʷəpse*, containing *śḥe* ‘head,’ *kʷe* ‘space, middle,’ and either *-ps(e)* ‘thread, string’ or *pse* ‘soul, life

principle' (there is also a variant with a different third element with unknown semantics), Kolyma Yukaghir *jo:n-qodo* 'head-lying,' Santiago Mexquititlan Otomí *ñäxumo* /ñä-ximo/ 'head-calabash,' Pawnee *paks-kiraar-uˀ* 'head-liquid-NOM,' Chayahuita *motoro'* /moto'-ro'/ 'head-CLASS.PILE,' Cubeo *jipo-jia* 'head-CLASS.RIVER.LIKE.OBJECT,' Jarawara *tati afone/tati afone* 'head soft.core,' Rama *kíng kás* 'head meat,' Tehuelche *č'eter ʔašk'en*, containing *č'eter ~ č'eʔter ~ četer* 'head' and *-ašk'en* 'interior,' Malagasy *atidòha*, analyzable as /àty-n-lòha/ 'liver/center-GEN-head,' and Rotuman *'uat pạr-pạru* 'head RED-mix.or.knead.with.hands.' Further, in both Greek and Welsh, terms for 'brain' are (etymologically) connected to 'head,' amended by a prefix meaning 'in' (Buck 1949: 215). Semianalyzable terms where the identifiable constituent is 'head' are found in Kaluli, Carrier, Haida, the Cuisnahuat dialect of Pipil, Carib, and Cavineña.

Another class of terms for 'brain' bears associations to cognitive processes. Efik, Laz, Lesser Antillean Creole French, Nuuchahnulth, San Lucas Quiaviní Zapotec, and White Hmong colexify 'brain' with 'mind' or like meanings, Guaraní colexifies 'brain' with 'understanding, judgment' and 'talent' and Basque with 'intelligence,' while Kashaya *hoʔɬo* contains *-ᵒɬo* 'to think' and Miskito *won sinska* contains *sins* 'intelligence.' A further language where the word for 'brain' appears to be related to cognitive abilities is Toba, but the precise structure remains unclear.

Further, Ket and Lake Miwok colexify 'brain' with 'semen.' An association restricted in the languages of the sample to the Barbacoan languages Cayapa and Tsafiki is that with 'faeces' (*mishpe* /mishu-pe/ 'head-excrement' and *fu-pe* 'hair-excrement' respectively).

Other associations include: Anggor *efu* is glossed also as "bee (sweat)," and Mali *genaing* also means 'phlegm, mucus,' while the Meyah term *ofóita* also means 'pulp.' Sahu *nyinyiala* appears to be derived by reduplication from *nyiala* 'canarium tree.' Badaga *mu:ḷe* also means 'skull,' Ket *doˀŋ* also 'three,' and Khalkha *tariki(n) ~ taraki(n)* also 'occiput.' The Nez Perce term *yexyeqíˀ-ɬes* is analyzable as 'tan-INSTR.' The underlying verb means more specifically "to put in a solution of brain as a step in tanning hide." Cashinahua *mapu* also means 'ashes,' 'soap,' the '"head" of an axe,' and is also used with reference to that part of a rudder where it is held. Guaraní *apy-tu'û* appears to be analyzable as 'extreme.point-softness,' and Ancash Quechua *tuqshu ~ toqshu* can also mean 'stupid.' Kapingamarangi *ngogo* also means 'egg' and 'zero,' Sedang *ngoa* might contain *ngo* 'mountain,' and Tetun *kakutak* is derived by the nominalizing circumfix *ka-...-k* from the verb *kuta* 'to smear.'

105. The Breast

Representation: 98%
Motivated: 57.8%
Thereof Analyzable: 5.2% Thereof Colexifying: 53%
Thereof by Contiguity: 32.5% Thereof by Similarity: 5.6%
Recurrent associated meanings: milk, udder/teat, nipple/teat, suck, chest, heart, mammary gland, Burton's legless lizard, bay

The most common association for the 'breast' is that with '(mother's) milk,' occurring in as many as 49 sampled languages without clear areal tendencies, though the virtual absence of the pattern in Eurasia is notable. The relevant languages are Ngambay, Noni, Yoruba, Anggor, Baruya, Buin, Burarra, Gurindji, Kyaka, Lavukaleve, Ngaanyatjarra, Nunggubuyu, Toaripi (colexifying also 'scrotum' and denoting a particular breast-shaped shellfish), Sahu, Kosarek Yale, Basque, Kolyma Yukaghir, Carrier, Upper Chehalis, Cheyenne, Haida, Kiowa (colexifying also 'to flow, melt'), Lake Miwok, Nez Perce, Nuuchahnulth, Oneida, Santiago Mexquititlan Otomí, Quileute, Xicotepec de Juárez Totonac, Wappo, Wintu, Yuki, Central Yup'ik (Yukon and Norton Sound dialects), San Lucas Quiaviní Zapotec, Aguaruna, Aymara, Cashinahua, Huambisa, Kaingang, Macaguán, Bislama, Hani, Hawaiian, Lenakel, White Hmong, Rotuman, Samoan, Takia, and Tetun; Hausa colexifies 'breast' with 'sour milk' specifically, in Swahili, the association is realized by noun class alternation, and in Kashaya *ši?do* is also the archaic term for 'milk,' while nowadays *molokko*, a loanword from Russian, is used.

The Haida term *tl'tn-ʔu* colexifying 'breast' and 'milk' is analyzable as 'suck-INSTR,' and so is the relevant Kolyma Yukaghir term. A derived term is also found in Rotokas, Sora (here, the term can dialectally also refer to the 'mother' herself), Upper Chehalis, and Tetun (and a semianalyzable one in Chukchi), while Muna, Basque, Itzaj, Lesser Antillean Creole French, Pawnee, Kapingamarangi, Samoan, and Bislama have terms directly colexifying nominal 'milk' with verbal 'to suck,' sometimes among other meanings. In Fijian *sucu* is 'to be born, suck the breast, birth, milk,' with milk also expressed by *wainisucu* (*wai* 'water'), and 'breast' is *sucuna* (*-na* is a possessive marker).

The Nez Perce and Lavukaleve terms colexifing 'milk' and 'breast' at the same time also colexify 'nipple' and/or 'teat' (see Buck 1949: 248 for Indo-European evidence), and this is a pattern also found in Khoekhoe (here with 'nipple of man' specifically), Dongolese Nubian, Swahili, Yoruba, Northeastern Tasmanian, Badaga, Basque, Khalkha, Kolyma Yukaghir, Ineseño Chumash, Haida, San Mateo del Mar Huave, Itzaj, Lake Miwok, Nuuchahnulth, Pawnee, Pipil, Wintu, Aguaruna, Arabela, Aymara, Bora, Cavineña, Chayahuita, Embera (where the meanings are associated with different genders), Guaraní, Lengua, Miskito, Ancash Quechua, Toba, Yanomámi, Hawaiian, and Sedang, which colexifies also "to pound, to beat" (34 languages). Hausa and Kwoma colexify 'mammary gland.'

Moreover, Cubeo *opebo* 'breast' might contain *ope* 'nipple,' but this cannot be determined with certainty on the basis of the consulted source. In many sampled languages, terms for 'breast' are also capable of referring to the homologous structure in animals, namely the 'udder' or 'teat.' This is found in Efik, Hausa, Khoekhoe, Dongolese Nubian, Yoruba, Kyaka, Toaripi, Badaga, Japanese (only in a term restricted to child language), Khalkha, Sora, Kildin Saami, Haida, San Mateo del Mar Huave, Itzaj, Lake Miwok, Nez Perce, Nuuchahnulth, Pawnee, Wintu, Yaqui, Aguaruna, Arabela, Aymara, Bora, Cavineña, Chayahuita, Embera (where the meanings are associated with different genders), Guaraní, Hupda, Lengua, Miskito, Pipil, Ancash and Imbabura Quechua, Toba, Yanomámi, Hawaiian, Malagasy, Mandarin, White Hmong, Rotuman, Takia, and Yay (44 languages). Hawaiian and Mandarin have optional complex terms on the basis of a term colexifying 'milk' and

'breast' that exhibit this pattern: *puʻu-waiū* 'protuberance-milk/breast' (which also colexifies 'wet, moist') and *ru3-fang2* 'milk/breast/drink-room' respectively.

Moreover, like English and other European languages, fifteen sampled languages, Sentani, Badaga, Greek, Khalkha, Welsh, Blackfoot, Cahuilla, Nuuchahnulth, Wappo, Bora, Embera (with gender differentiating between the meanings), Rama, Fijian, and Vietnamese, colexify 'breast' with 'chest' (Badaga also with 'arm'), and four sampled languages, Katcha, Highland Chontal, Kaingang, and Great Andamanese with 'heart' (Kolyma Yukaghir also with 'soul' and 'breath').

Further, in Burarra and Nunggubuyu, relevant terms for 'breast' and 'milk' are also used to refer to a species of lizard (Burton's legless lizard, *Lialis burtonis*), and in Basque and Greek, relevant terms can also refer to a 'bay.' Both associations are likely areal.

Other associations include: Efik *ëba'* is derived from *iba'* meaning 'two' among other things. Hausa *nono* also denotes "the fins below the head of a fish" as well as "a cluster of fruit," and *mama* is a "title and position among female girls" and also means "throwing pestle into the air each stroke when pounding." Dongolese Nubian *érti* has a homonym (?) meaning 'dirt, dirty,' Rendille *náhas* also means 'pump, waterpipe,' Yoruba *ọmú* also denotes the "instrument used in weaving to divide the woof," and Muna *titi* also means 'stalactite' inter alia. Sentani *nimə* also means 'ripe,' and Badaga *boːsi* also denotes the "act of fondling breasts" as well as "tumbler, glass" in the Kunde dialect, and *mai* can also refer to a "deep location." Abzakh Adyghe *ǯə* also means 'to throw, hurl,' Basque *golko* can also refer to the 'stomach, guts' and *bular* also means 'courage' inter alia. Japanese *chichi* also means 'father' (the meanings being distinct in writing). Khalkha *cegezi(n)* also may refer to the "memory as a faculty," and *elige*, figuratively, also means 'liver,' 'belly,' and "blood relatives," while Welsh *bron* is extended to 'breast of hill.' Lesser Antillean Creole French *sen* also means 'holy' and 'sane' inter alia, and Lake Miwok *łéele* also 'to be slicing meat.' Nuuchahnulth *ʔinma* also denotes "Nob Point, where white powder seems to run out of a breast-shaped rock," and Pawnee colexifies 'breast' with 'body, corpse' inter alia. Wintu *ʔEm* can also mean "hold pectorally, carry something in the arms, embrace." Copainalá Zoque *cucpac* contains *pac* 'bone,' and *tzu'tzi* also denotes the "toothing of a stamp mill," while Bororo colexifies 'breast' with 'bud/shoot of plants' and Wayampi with 'stamen' of a flower. Macaguán *-apúchipar* ~ *-atbu̶chipar* contains *-chipár* 'finger,' Miskito *lama* also means 'present, benefit' and 'presence, proximity.' Piro *sta* colexifies 'front,' and there is also a verb *-sta-* 'to cut, tear, break.' Bislama *titi* and *susu* also mean "unweaned," as said of a child, and the former term also means 'to practice fellatio,' while Fijian *sere* also means 'to sit and sing' and 'to unloose, untie.' Hani *aqqul* also means 'sweet,' Hawaiian *ū* also 'moist, soaked' and 'to drip, drizzle, ooze' inter alia, and *waiū* also denotes a 'wet-nurse.' Kapingamarangi *uu* also means 'to absorb' inter alia. Finally, Rotuman *susu* also means 'to sew,' Samoan *mau*, a polite term for 'woman's breast,' also 'to keep, retain' and 'to live, dwell' inter alia, and Takia *su* also 'shoe' (with the additional sense < Engl. *shoe*).

106. The Buttocks

Representation: 84%
Motivated: 55.4%
Thereof Analyzable: 17.5% Thereof Colexifying: 37.4%
Thereof by Contiguity: 22.4% Thereof by Similarity: 17.9%
Recurrent associated meanings: bottom/base, back/behind, hip/haunch, anus, rump, end, tail, seat, cheek, faeces, stern of boat, thigh, hole, reason, root, sit

The most frequent association for the 'buttocks' is that with 'bottom,' occurring in 26 sampled languages (and also in English). There is a grammaticalization path said to be particulary common in African languages from 'buttocks' to 'down' (Heine and Kuteva 2002: 63); however, it is hard to impossible to tease apart these cases from those where relevant terms are glossed as 'bottom' as a slightly euphemistic alternative for 'buttocks' themselves. All cases are reported here in spite of the fact that some may be spurious, though it is notable that in many cases the extension is said to be explicitly to 'bottom of a vessel,' so that there are clearly also genuine cases. The association with 'bottom' or 'base' is attested by colexification in Buli, Efik, Hausa, Rendille, Buin, Burarra, Kwoma, Mali, Muna, Ngaanyatjarra, Kosarek Yale, Badaga (by an archaic term), Basque, Sora, Lesser Antillean Creole French, San Lucas Quiaviní Zapotec, Aymara, Carib, Guaraní, Ancash Quechua (where the relevant term is also an insult), Great Andamanese, Kapingamarangi, Lenakel, Malagasy, Rotuman, and Bislama. Tetun has *kidun-tahan* 'bottom-flap,' and Meyah *oskú ofogú* 'bottom flesh' (a semianalyzable term with 'flesh' is also found in Kyaka). Note also Kosarek Yale *di kumkum* 'faeces neck/base' and *dibomaak* 'faeces-front.end/lower.part/bottom;' semianalyzable terms with 'bottom' are furthermore found in Sko (where the putative constituent means "bottom of a four-legged animal" more specifically), Waris, and Wappo. Another cross-linguistic association related to a grammaticalization path (Heine and Kuteva 2002: 62) is that with 'behind.' Here, the same cautionary remarks made above for 'bottom' apply. 'Back' and/or 'behind' or 'hinder part' are colexified in Buli, Ngambay, Dongolese Nubian, Rendille, Badaga, Basque, Sora, Oneida, Carib, Huambisa, Rama, Fijian, Hawaiian (where the relevant term bears other related meanings as well), and Rotuman. Moreover, Chickasaw has *im-ą́shaka'* 'DAT-back,' and Nuuchahnulth *ʔamasʔakƛ̓i*, analyzable as /ʔam-as-'akƛ̓i/ 'in.the.centre-approach-behind' (a complex term featuring an element meaning 'behind' may have been present in Middle-Eastern Tasmanian). Somewhat similarly, 'buttocks' betrays an association with 'end' by colexification in Efik, Dongolese Nubian, Basque, San Lucas Quiaviní Zapotec, and Jarawara, while Abzakh Adyghe has *čʔentʔəʔ°* /čʔe-n-tʔ(e)-ʔ°(e)/ 'end-RELAT/EPEN-poke.out-peak' and Kosarek Yale the aforementioned term *di-bomaak* 'faeces-front.end/lower.part/bottom.'

There are also contiguity-based associations with adjacent parts of the body: Dongolese Nubian, Rotokas, Toaripi, Badaga, Sora, Cheyenne, Lakhota, Pipil, Aguaruna, Bora, Jarawara, Hawaiian (the relevant term *'ōkole* may be derived from *kole* 'raw, inflamed, red,' there is another Hawaiian term with a similar structure containing *kole*), and Bislama colexify 'buttocks' with 'anus,' and Baruya features a complex term with the constituents

'anus' and 'hill.' 'Buttocks' is colexified with 'hip' or 'haunch' in Dongolese Nubian, Kwoma, Kosarek Yale, Japanese, Kildin Saami (by the analyzable term *čuarr-piell'* 'pelvic-half'), Welsh, Upper Chehalis (by a term containing a constituent meaning 'elbow;' the gloss 'hips' is put in parentheses in the source), Yuki, Central Yup'ik (Nunivak Island dialect), Cashinahua, Piro, Hawaiian, Bwe Karen (where the relevant term contains *-ká* 'hind part'), and Sedang, in which latter the relevant term *kɤtei* is also the name of a gray bird. Japanese further has the analyzable term *den-bu* 'hip-part,' Bororo *etawu oto* 'hip point,' and Mandarin *pi4-gu* 'fart-upper.thigh/hip/member.' Mandarin is not the only language betraying an association with 'thigh': Ngambay and Kildin Saami (by the same analyzable term just mentioned) colexify 'thigh' with 'buttocks,' and Chickasaw has *iyyobi' ishto-ka'* 'thigh be.big-NMLZ.' Toaripi, Kosarek Yale, Khalkha, Biloxi, Cheyenne, Kiowa, Lesser Antillean Creole French, Pawnee, Embera (by the term *ãndáuboregéa*, analyzable as /ãndáu-booregéa/ 'rear/hind-fat'), Miskito, Wayampi, and Fijian colexify 'buttocks' with 'rump.' There is also one association with another body-part that seems metaphorical in nature, namely that with 'cheek.' Basque and Pawnee colexifies these meanings, while Muna has *bhaga-no koro* 'cheek-POSS buttock,' Kashaya *sili qaṗa* 'rear jowls,' and Wichí *towejch'alus* contains *wej* 'queue' and *ch'alu* 'cheek.'

As discussed above, Kosarek Yale has a complex term where one constituent is 'faeces;' this association is also found by colexification in Takia and by overt terms also in Bora (*námehéyu* /name-héyu/ 'faeces-CL.hole'), Jarawara (*joto-hoti* 'faeces-hole'), and Lenakel *nɨmwa-nɨsii-* 'covering-excrement-;' for the association with 'hole,' note also Kolyma Yukaghir *ńerčəd-aŋil'* 'bad-hole.'

There are also metaphorical extensions of 'buttocks,' sometimes to more abstract meanings: Yoruba and Mali colexify it with 'reason,' Hausa and Yoruba with 'root' (Hausa also with 'foundation'), and Haida, Guaraní, Miskito, and Hawaiian with 'stern of boat.' Moreover, Rendille, Basque, and Kolyma Yukaghir colexify 'buttocks' with 'tail,' while Rama has *túkakás*, analyzable as /tuk-kás/ 'tail/end meat,' White Hmong *pob-tw* 'ball-tail,' and in Khoekhoe, the same root is used with alternating nominal designants to convey the two meanings. Similarly, Ngaanyatjarra colexifies 'tail of insect' specifically. Finally, Buli, Basque, Rama, Samoan, and Yay colexifies 'buttocks' with 'seat' (Samoan also with 'dwelling, residence' and 'station;' the meaning 'buttocks' is polite). Similarly, Yuki and Samoan feature terms derived from verbs meaning 'to sit.'

Other associations include: Buli colexifies 'buttocks' also with "outside, outer surface," as said of pots, and, by a different term, with 'descendants,' in particular 'grandchildren.' Ngambay *bàgìrì* also denotes a "sieve, type of basket to remove husk from sesame, millet and gourd seeds." Yoruba colexifies 'buttocks' with 'waist.' Buin *koku* also means 'valley' and 'point of a banana rope,' while Muna *koro* also denotes a type of rooster, and One *you* also means 'call.' Abzakh Adyghe *čʲentʲəʔ°* also means 'kidneys' and Badaga *kuṇḍe* also 'olive' inter alia. Sora colexifies 'buttocks' with 'rectum,' and Chickasaw *ishkish* is also used to refer to the sexual organs in general. The Haida term *ruda* is also used with the meanings "the outside part of certain body parts," "that part of an island nearest to other land" and "side of house." Kiliwa *ʔuuw-h-paʔ* is analyzable as 'vagina-3-set.round.object.down.' The Kiowa root *t̂ei-* in *t̂ei-dl* 'buttocks' also means 'calf' when used

with the noun postfix *-p*. Oneida *ohná·tsha?* also means 'small stool,' Tuscarora *uhnę́·neh* and Guaraní *topyta* also 'trunk' (San Lucas Quiaviní Zapotec colexifies 'trunk of car' specifically), and the Wintu term *werem ɬah* contains *ɬah* 'nearby.' Miskito *pnata* also means 'croup.' Great Andamanese *ardama* appears to contain *dama* 'lean' – in contrast, Hani *daoqmeil* seems to contain *meil*, meaning 'fat, obese' inter alia. Hawaiian *lemu* is also used with the meaning 'slow-moving, sluggish, lagging,' and *ʻōkole* is also the name of a sea creature. Rotuman *muri* also means "young, not having reached maturity."

107. The Calf

Representation: 79%
Motivated: 43.4%
Thereof Analyzable: 34.0% Thereof Colexifying: 10.5%
Thereof by Contiguity: 13.0% Thereof by Similarity: 28.5%
Recurrent associated meanings: leg/foot, belly/stomach, muscle, shin, egg, knee, pregnant/pregnancy, cheek, scrotum, fruit, flesh

Motivated terms for the 'calf' are more often morphologically complex than colexifying, and more often driven by metaphor than by metonymy. In most languages with complex terms, one of the constituents means either 'leg' and/or 'foot.' In ten sampled languages, 'belly' or 'stomach' is the meaning of the second constituent, as in Wichí *tot-kolo-ts'e* 'POSS.INDEF-leg-paunch.' Such terms are also attested in Efik, Sko, Toba, Malagasy, Mandarin, White Hmong, and Tetun. In addition Welsh colexifies 'calf' with 'belly, womb' and by another term with 'belly, stomach' (and has additional redundant complex terms), and Cashinahua has *bipustu*, where *pustu* is 'stomach' and *bi* 'mosquito' and 'anteater,' alongside *bi-tuxtu*, where *tuxtu* is 'rounding.' In Pawnee, furthermore, 'calf' is *kaac-karaar-u?* 'gray-belly-NOM.' Compare Sadovszky (1973) for discussion of this association and Matisoff (2004: 358) for further examples; Sadovszky also demonstrates a widespread association of 'fish' or more specifically 'fish roe' with the 'leg' and the 'calf of the leg' in Eurasia more specifically.

Similarly, Wayampi has *tɨmã-puluʔa* 'leg-be.pregnant,' and this metaphorical transfer is realized by colexification with 'visibly pregnant' in Ngaanyatjarra, where the latter meaning is register-specific. Swahili has *shavu la mguu* 'cheek of leg,' and Hawaiian colexifies 'calf' with 'cheek' directly, alongside other meanings such as a container made from a long gourd. Biloxi has *yupkĕ́-iⁿtí* 'leg-egg.' An analyzable term of the lexical kind is also found in Haida, and one of the derived kind in Burarra. In Takia, the 'calf' is *ŋe-n labe-n* 'leg-3SG scrotum-3SG,' while Hawaiian colexifies 'calf' and 'scrotum' (among other meanings). Yaqui has *woktomam* /wokim-tomam/ 'feet-muscles;' such a term (with 'leg' rather than 'feet') is also featured in Kashaya (here, *ʔahpʰeṭ* curiously colexifies 'muscle' with 'mussel,' in particular *Mytilus californianus*), while Khalkha, Cheyenne and Tehuelche colexify 'calf' with 'muscle' generally; moreover, Badaga *moṇṇe kaṇḍa* is analyzable as 'lower muscle.' There are also many other complex terms where one of the constituents is 'foot' and/or 'leg': Yoruba has *isu-ẹsẹ̀* 'yam-leg,' Basque *zango-sagar* 'leg-apple,' Upper Chehalis *sáẇtiyq* contains *ʔáw-* 'behind' and *=iyq* 'foot/leg,' Chickasaw has *iyyintakaaliˈ*, analyza-

ble as /iyyi' im-takaali-'/ 'leg DAT-be.hung.up.on-NMLZ,' San Mateo del Mar Huave *mi-pemb oleajaran* 'AL.POSS-water.pitcher/gourd foot/leg,' Kiliwa *miy=ha?* 'leg-face' (marked with a question mark in the consulted source), Lesser Antillean Creole French *gwat janm* 'grater leg,' Santiago Mexquititlan Otomí *dot'u̲wa*, questionably analyzable as /tot'i-wa/ 'fold/bend-foot/leg,' Yana *ʒaʔp̣gulgadu* /ʒaʔp̣gul-gaadu/ 'full/be.filled-leg,' Carib *-ɨsai-punu* '-leg-chair,' Guaraní *tetyma ro'o* /tetyma to'o/ 'leg flesh' (for which compare Kwoma *hapa omu* 'bone flesh/seed/fruit,' for which in turn compare Bwe Karen *kha-dé-θɛ* 'leg-narrowest.part-fruit'), Rama *kát-aring* 'foot-shit,' Wayampi *tɨmãkuʔa* 'leg-middle,' Fijian *temo ni yava* 'thick.part.of.limb POSS leg,' Rotuman *parpar ne lā* /parpara ne lā/ 'soft ART.PL foot/leg,' Samoan *ate-vae* 'liver-foot/leg,' Vietnamese *bắp chân* 'shaft leg,' and Yay *ʔaay[1] ka[1] (raay[4])* 'goiter leg (mark).' Finally, Comanche has *taʔwiitsa̲*, seemingly analyzable as /ta-wiitsa̲/ 'foot-leg,' there is a semianalyzable term where the identifiable constituent means 'leg' in Bakueri, where it means 'back side of leg' specifically in Cayapa, and a semianalyzable term involving a lexical affix for 'foot, toe, leg from knee down' in Upper Chehalis.

Greek, Quileute, Cubeo, Miskito, and Imbabura Quechua colexify 'leg' with 'calf' directly (and Hani with 'lower leg below knee' more specifically), while the referent of the relevant Pipil term varies semantically between 'leg' in the Cuisnahuat dialect and 'calf of leg' in the Santo Domingo de Guzmán dialect. Semianalyzable terms are furthermore found in Bakueri, Buli, Mbum, Anggor (the meaning of the entire term is considered unsure by lexicographers; the unknown element is *mbosɨmundɨ*, compare *mbusɨmondɨ* 'middle sibling, thorax of spider'?), Kemtuik, Ngaanyatjarra, Sentani, Abzakh Adyghe, Badaga, Sora, Chickasaw, Highland Chontal, Oneida, Embera, and Lenakel.

Further, Badaga, Chickasaw, and Sedang colexify 'calf' with 'shin' (Sedang also with 'to pick vegetables,' 'to transplant rice'); analogously, Japanese has *fukura-hagi* 'swell-shin,' and Great Andamanese has *abchâltadama* containing *abchâlta* 'shin' and *dama* 'lean' and *tâl'ârdama* containing *tâ* 'bone' alongside the element meaning 'lean.' A semianalyzable term with 'shin' is also found in Carrier. Finally, Kosarek Yale and Badaga colexify 'calf' with 'knee,' Aguaruna with 'hollow of the knee,' and Yir Yoront has *nhal-kar* 'back/inside.of.knee-like.'

Other associations include: Hausa *sha ra'ba* is analyzable as 'drink dew' (with *ra'ba* however also meaning 'to crouch near something' inter alia). The whole term can also refer to a "metal ornamental point to a sword or knife sheath." The Khoekhoe root *lkhoe* yields the meanings 'calf (muscle)' and "meat of shank, knuckle" depending on the nominal designant suffixed. Muna *lambi*, as a verb, means 'to hang,' and Kosarek Yale *buding* is also the name of a tree frog species and a variety of taro. Basque (Lower Navarrese dialect) *aztal* also means 'heel,' and Kolyma Yukaghir *činčəde* (containing *činčə* 'leg muscles') may also refer to the 'back part of boots.' The Kiowa root *t̂ei-* in *t̂ei-p* 'calf' also yields the meaning 'buttocks' when used with the noun postfix *-dl*, and Wintu *ċuhcir* may be related to *ċOy* 'to sprawl, stretch.' Xicotepec de Juárez Totonac *i'xchāpa'ta'kāt* contains *cha'pá'* 'pole,' and Central Yup'ik *nakacugnaq* consists of *nakacuk* 'bladder' and the postbase (see § 4.4.2.) *-naq* 'one like,' while Copainalá Zoque *po'cpo'c* is a reduplication of *po'c* 'knot, trunk of tree, blain.' Chayahuita *pa'o ~ pa'huë* also means 'butt of a gun,' and Yanomámi *shiãpɨ* also means

'thick part of an axe that doesn't cut.' For Ancash Quechua *pinkuullu*, compare *pinkullu* 'slim flute.' Tetun *kloor* may also refer to a 'hamstring, hock,' 'footstring,' or a 'trace.'

108. The Cheek

Representation: 93%
Motivated: 26.3%
Thereof Analyzable: 10.2% Thereof Colexifying: 16.5%
Thereof by Contiguity: 20.3% Thereof by Similarity: 4.2%
Recurrent associated meanings: face, jaw, mouth, cover, temple, chin, buttocks, edge, meat/flesh

'Cheek' (ignoring additional glosses like 'side of face') is often colexified with 'jaw,' namely in Buli, Yoruba, Gurindji, Yir Yoront (by the term *wal-kur* 'cheek/temple-slashing.sword'), Welsh, Lesser Antillean Creole French, Wintu, Abipón, Miskito, Kapingamarangi (by the term *gau wae*, which is seemingly analyzable as 'side foot/leg'), and Manange (see Buck 1949: 220-221 for the strong diachronic ties between terms for 'cheek,' 'jaw' and 'chin' in Indo-European). In Basque there is dialectal variation of the meaning of *baraila*, meaning 'jaw' alongside 'revelry, fray' in the Biscay dialect and 'cheek' in the Zuberoan dialect; further, for Hani *baqba*, compare *baqxoq* 'jaw.' Common are also associations with 'face' (and sometimes due to further colexification also with 'front'), by colexification in North-eastern Tasmanian, Abzakh Adyghe (by the analyzable terms *nek$^{\circ ?}$əś$^{\circ}$e*, containing *ne* 'mouth' and *ś$^{\circ}$e* 'skin' and *nek$^{\circ ?}$əṡḥe*, containing *ne* 'mouth' and *ṡḥe* 'upper part'), Chickasaw, Lesser Antillean Creole French (where also 'figure, form' and other meanings are colexified), Nez Perce, Wintu, Aguaruna, Bororo, Carib, Hupda, Ancash Quechua, Tehuelche, and Sedang. In addition, Abzakh Adyghe has *nepaś$^{\circ}$e* /nape-ś$^{\circ}$e/ 'face-skin' (a semianalyzable term with 'skin' is found also in One), Kiowa has *t̂ou-p̂ą'ęgyH* 'face-in.the.middle.of,' Guaraní *tova-yke* 'face-side,' Imbabura Quechua *ñawi chichu* 'face pregnant,' and Rama *mngut kás ~ múngut kás*, containing *ngút* 'face' and *kás* 'meat.' A term with identical structure is found in Maxakalí, and for Carrier, the source remarks: "the Carrier for cheek seems to mean face-profile, or nearly so" [sic!]. Semianalyzable terms for 'cheek' where the identifiable constituent is 'face' are furthermore found in Xicotepec de Juárez Totonac, Yana, and Tsafiki. Alongside the Abzakh Adyghe term just mentioned, there are also associations between 'cheek' and 'mouth' in other languages (in diachrony, Romance terms for 'mouth,' such as Spanish *boca*, go back to Latin *bucca* 'cheek,' see also Buck 1949: 221 on the cognacy of Germanic terms for 'cheek' with Avestan 'mouth'): they are colexified in Burarra (also with 'lips;' the term has broad reference to the 'mouth area' in general), and analyzable terms are found in Mbum (*fǎ-háù* 'sack-mouth;' similarly, 'bag' and 'cheek' are colexified in Ngambay), Sora (*kub-ma·b-'tam-ən* 'unite/be.covered-??-mouth-N.SFX; similarly, Efik has a derived term from a verb meaning 'to cover' and Rotuman colexifies 'cheek' and '(to) cover' inter alia), and Kiliwa (*yuw=ha?=kw-cas* 'eye=mouth=WH-??,' with *yuw=ha?* meaning 'face;' there is a semianalyzable term where the identifiable constituent means 'eye' in Abzakh Adyghe). Dialectally, the meaning of Yana *bal(la)* varies between 'cheek' and 'mouth.' Semianalyzable terms with 'mouth' are

also featured in Baruya and Kiliwa. Dongolese Nubian and Yir Yoront colexify 'cheek' with 'temple' (Yir Yoront alongside 'ear' and 'sleep'), and Burarra and perhaps Abipón with 'chin.' Basque and Pawnee colexify 'cheek' with 'buttocks,' Katcha colexifies 'cheek' with 'edge,' and Copainalá Zoque has *aca-pac* 'edge-bone' (an element meaning 'bone' is also etymologically recoverable in the Ket term for 'cheek').

Other associations include: Lavukaleve *hou* is also used with the meaning 'smoke,' while Muna *bhaga* also means 'molar tooth' (there is a semianalyzable term containing an element meaning 'tooth' in Berik; compare also Nunggubuyu *ṟamara* 'cheek' and *ṟa:* 'tooth'). Toaripi *heva* also means 'gill.' Sko *òebi* appears to be analyzable as /òe-bí/ 'penis-shell/floor' and also means 'testicle,' while Tuscarora *unhú?weh* also means 'areola of nipple.' Arabela *sacomara* 'cheek pouch' appears to consist of *saco* 'hollow fruit' and *mara* 'swamp.' Jarawara *abate/ebete* also means 'tongue,' and Tehuelche *q'ape-n-k'en* is analyzable as 'be.red-??-LOC.' Fijian *balu* also denotes the 'sides of the head of a club,' while Hani *baqba* may be related to *baq* 'direction,' 'thin' and *ba* 'light in color, white.' Hawaiian colexifies 'cheek' and 'calf of leg,' *papālina* ~ *pāpālina* is perhaps analyzable as /papa-lina/ 'flat.surface-soft' (both *papā* and *pāpā* exist as well, but do not have meanings standing in any obvious relationship with 'cheek'), and Bwe Karen *-ɓɔ* also means 'packet, parcel' and "creeper with a bitter shoot."

109. The Chin

Representation: 90%
Motivated: 44.2%
Thereof Analyzable: 12.8% Thereof Colexifying: 33.0%
Thereof by Contiguity: 40.5% Thereof by Similarity: 3.8%
Recurrent associated meanings: jaw, beard, mouth, lower/below, bone/skeleton, cheek

'Chin' is very commonly colexified with '(lower) jaw' cross-linguistically (see Buck 1949: 220-221 for the strong diachronic ties between terms for 'cheek,' 'jaw,' and 'chin' in Indo-European). This is found in as many as 44 languages, namely Yoruba, Baruya (by the analyzable term *maanag̈inya* /maanga-yag̈inya/ 'mouth-bone'), Kaluli, Kwoma (by the term *teekibi*, perhaps containing *teeki* 'to string, tense' and *bi* 'point'), Kyaka, Lavukaleve, Mali, One, Rotokas, Southeastern Tasmanian, Toaripi, Kosarek Yale, Abzakh Adyghe (by the analyzable term *że-pqʔ* 'mouth-skeleton'), Greek, Japanese, Ket (by a semianalyzable term containing *ìt* 'tooth,' an element with that meaning is also found in the Berik, Embera and Kaingang terms), Khalkha, Nivkh, Welsh, Kolyma Yukaghir, Blackfoot, Upper Chehalis, Cheyenne, idiolectally in Chickasaw, Itzaj, Kiliwa (by a semianalyzable term containing *haʔ* 'mouth'), Nez Perce, Nuuchahnulth (by the analyzable term *hiiṅiiƛaksuɬ* /hiiṅiiƛa-ʔaksuɬ/ 'lower/below-at.the.mouth/at.the.lips' which also denotes the 'lower lip'), Pipil (by the analyzable term *-te:n-tsi:ka-w* '-mouth-ant-POSS'), Arabela, Aymara, Cashinahua, Embera (where the meanings are associated with different genders), Guaraní, Jarawara, Macaguán (by a semianalyzable term containing *-bʉk* 'heel'), Rama (by a semianalyzable term containing an element meaning 'tree, foot' and colexifying 'grater'), Tehuelche, Yanomámi, Mandarin, White Hmong, Samoan (by the analyzable term *ʻau-vae*

'CLASS.long.or.narrow.things-foot/leg' which also means 'foot of hill), Takia (by the analyzable term *awa-n to-n* 'mouth-3SG arm-3SG'), Tetun, and Yay; a semianalyzable term is found in Mandarin. Complex terms for 'chin' on the basis of 'jaw' are Biloxi *yatka´ psû^{n}tí´ ~ i´yatka´ psû^{n}tí´ ~ nyatka´ psû^{n}tí´* /yatka´ pû´tsa/ 'jaw sharp,' Cayapa *tejmashmutu* /tejmashi-mutu/ 'jaw-corner,' and Chayahuita *hui'nin cohuiratë'* 'child/prow jaw.' Associations with 'beard' are also quite common, also in Indo-European (Buck 1949: 224). 'Beard' and 'chin' are colexified in Dongolese Nubian (which also colexifies 'whiskers'), Gurindji, Ngaanyatjarra, Tasmanian (Northeastern, Middle Eastern, and Western), Badaga, Embera, Fijian, Rotuman, and Tetun. Further, Kanuri has *njìtì-rám* 'beard-place.of,' in Swahili, the term for 'chin' consists of that for 'beard' and a noun class prefix, and semianalyzable terms are found in Buli and Sentani (note that Ancient Greek had a term literally translated by Buck 1949: 224 as 'that which has a beard on it' and Russian one literally meaning 'what is under the beard,' there is a semianalyzable term where the identifiable constituent means 'under' inter alia in Wintu). Bora has *újcawa*, with the same root possibly also occurring in the word for 'beard,' but segmentation is uncertain. As already seen in some of the above terms, there are also associations between 'chin' and 'mouth:' Burarra has a general term referring to the mouth area in general, including the 'chin,' and complex terms are found in Baruya (*maanagɨnya* /maanga-yagɨnya/ 'mouth-bone'), Abzakh Adyghe (*żepqʔ* /że-pqʔ(ə)/ 'mouth-skeleton'), Sora (*də'naŋkul'tam- ~ dənaŋkultam-*, containing *dənaŋ-* meaning 'obstruction, bar' inter alia and *'tam* 'mouth'), Kiowa (*beidl-t'ęį'm* 'lip/mouth-bone'), Nuuchahnulth (*hiiɬiiƛaksuɬ* /hiiɬiiƛa-ʔaksuɬ/ 'lower/below-at.the.mouth/at.the.lips'), Pipil (*te:n-tsi:ka-w* 'mouth-ant-POSS'), Lengua (*koning atang* 'below mouth'), and Mandarin (*xia4-ba*, with *xia4* meaning 'down, lower' and *ba* referring to the 'mouth'). Semianalyzable terms are found in Nunggubuyu and Kiliwa. For the associations with 'bone' above, note also that Carrier has a semianalyzable term where the identifiable constituent bears that meaning. Finally, Burarra and perhaps Abipón colexify 'chin' with 'cheek.'

Other associations include: Dadibi *penani* also means 'adultery, fornication,' One *sesu* also 'to squish,' Kosarek Yale colexifies 'chin' with "wattle of an agama," and Basque *kokots* also means 'snout,' 'calyx,' and 'mesh.' Carrier *-yěta* contains *ta* 'surface,' Central Yup'ik *tamlu ~ tamluq tamu-* 'to chew once,' and Jarawara *enekiri/enekiri* also means 'gill.' Macagúan *pipumáchipla* also can refer to the 'forehead,' Toba *lqa'* also is glossed as Spanish 'pera,' which is either 'pear,' 'bulb,' or 'signal horn' in English, while Wayampi *ɛnɨ-wa* is analyzable as 'saliva-eater.' Fijian *kumi* also denotes 'Tongan cloth,' and Bwe Karen *khɛ* also means 'to be bitter' inter alia. Hawaiian *'auwae* inter alia also denotes the "curved notch cut on the outer side of a post below the base of a tenon." Rotuman *kumkumu* also denotes a species of crab, and Tetun also means 'ambush' and "notch in wooden columns of buildings for the positioning of beams."

6.2.2.110. The Eyeball

Representation: 36%
Motivated: 86.9%
Thereof Analyzable: 81% Thereof Colexifying: 7.5%
Thereof by Contiguity: 17.5% Thereof by Similarity: 65.4%

Recurrent associated meanings: eye, seed, egg, pupil, child, fruit, grain, stone, ball, round object

Terms for 'eyeball' are overwhelmingly morphologically complex in the languages of the sample, and are often metaphor-based. Typically and obviously, 'eye' is one of the constituents, and the second one in metaphor-driven terms typically denotes a small round object. Ngaanyatjarra has *kuru yurniny(pa)* 'eye seed/grain,' and terms with an association with 'seed' are also featured in Buli, Efik (where 'seed' is *mpa'sïp ~ mkpasïp* /mp'kö-sïp/ 'thing-be.small'), Kwoma, Muna, Sahu, Kosarek Yale, Sora, Itzaj, Lesser Antillean Creole French, Wintu, Guaraní (though note that *ta'ŷi* 'seed' is a diminutive of *ta'y* 'son, clot'), Miskito, Bwe Karen, and Sedang, while Yay features a semianalyzable term where the identifiable constiuent is 'seed.' Kosarek Yale *wana* colexifies 'seed' with 'eye,' 'heart,' 'egg,' 'fruit,' and 'child' (compare § 6.2.3.1.), and, by virtue of this, the term for 'eyeball' *heing wana* also betrays an association with these meanings. An association with 'egg' is also found in a number of other sampled languages, as in Toba *l-'ai l-co'oue'* '3SG.POSS-eye 3SG.POSS-egg,' and also in Efik, Koyraboro Senni, Mbum, Yoruba, Sora, Wintu, Fijian, Kapingamarangi, and Lenakel. Wappo has *huċi·lél* /huči-lél/ 'eye-stone' (which also denotes 'hailstone'), and analogous complex terms without the additional colexified meaning exist in Koyraboro Senni and Kashaya (in Koyraboro Senni, the constituent meaning 'stone,' however, is said to assume the meaning 'egg' in compounds as well). Similarly, Kosarek Yale, Kwoma, and Miskito colexify 'seed' and 'fruit,' and therefore, in these languages, 'fruit' is (also) the source concept for the metaphorical transfer; the same situation obtains in Tetun. Further, Piro colexifies 'large round fruit' with 'ball,' and analogously to English *eyeball*, 'ball' is the meaning of the second constituent in Lake Miwok (*ṣút polóolo* 'eye ball'), in Kiliwa, and, by virtue of the language's pattern of colexification, in Piro. Similarly, Basque has *begi-globo* 'eye-globe.' Katcha has *bibala m-ɔe* 'child ??-eye,' and such terms are also found in Kosarek Yale, by virtue of the large semantic range of the constituent *wana* as discussed above, Welsh (colexifying 'boy, son' specifically rather than 'child' generally; this term is marked as being obsolete), Samoan, and, because of the internal structure of the term for 'seed' discussed above, also in Guaraní. A semianalyzable term with 'child' is featured in Yay, and similarly, Carrier has *tṣûten-pê-šta* 'child-wherewith-sit.in.' This pattern is likely intertwined with the association between the 'eyeball' and the 'pupil' (see also section 130); 'eyeball' and 'pupil' are colexified (sometimes by analyzable terms discussed elsewhere in this paragraph) in Hausa, Katcha, Yoruba, Welsh, Wintu (where the relevant term also denotes the 'iris' at the same time), and Guaraní.

Complex terms with 'round object' alongside 'eye' are featured in Chickasaw (*ishkin lobo'* 'eye round.and.firm.object) and Oneida, and Blackfoot has *ohkomapinssin* /ohkom-aapini-hsiN/ 'be.round-eye-NMLZ.' There is a complex term with 'white' in Kaingang (*kanẽ kupri* 'eye white'). Other complex terms of the lexical type with 'eye' being one of the constituents are Hausa *k'wayar ido* 'grain eye' (this association is also present in Ngaanyatjarra, as noted above, as well as Lesser Antillean Creole French. Similarly, Hani *miavneev* 'eyeball, eye' is derived from *miav* 'eye' by means of *neev*, the classifier for beans and grains), Mbum *máà-yâr* 'mother-eye,' Ngambay *dò-kèm* 'head/on-eye,' Kosarek Yale

heing ngei ‘eye ground/basis,’ Sora *'ranˑa-'mad-ən* ‘gold-eye-N.SFX’ (glossed in the source as ‘gem of the eye’ and marked with a question mark), Welsh *cannwyll y llygad* ‘candle DET eye,’ Cheyenne *menoo'éxané*, containing *mené* ‘berry’ and *éxané* ‘eye’ (gloss in the source: “globular (that is, berry).shaped-eye”), Lesser Antillean Creole French *koko zié* ‘coconut eye,’ Jarawara *noki kori/noko korone* ‘nakedness eye,’ Samoan *ʻiʻoi-mata* ‘tuber/corm-eye,’ and Bislama *mabol blong ae* and *pi blong ae* ‘marble POSS eye.’ Moreover, ‘eye’ and ‘eyeball’ are colexified directly in Buli, Koyraboro Senni, Toaripi, Yir Yoront, Laz, Nivkh, Jarawara, and Hani. In this case, the ‘eye’ as a physical object is probably the colexified meaning more precisely; semianalyzable terms for the ‘eyeball’ on the basis of ‘eye’ are found in Koyraboro Senni (where the second constituent is diachronically relatable to a word meaning ‘belly’), Yir Yoront (where the second constituent diachronically goes back to a word meaning ‘moon’), and Bislama.

Other associations are few in number: the Burarra term *munbarra* consists of the noun class prefix *mun-* and *barra* ‘rear end,’ Greek colexifies ‘bulb’ with ‘eyeball,’ and Nuuchahnulth *ƛ̓iskc̓iṅim* contains the verb root *ƛ̓isk-* ‘for the eyes to flash white.’ San Lucas Quiaviní Zapotec *bàa'ah* is also used with the meaning ‘iris of the eye,’ and Piro *tskata* also means ‘cliff’ or ‘bank.’ Hawaiian *pipi* also denotes the ‘Hawaiian Pearl Oyster’ inter alia, and for *ʻōnohi*, compare *nohi* ‘bright-colored, vivid’ and the ‘similitude’ prefix *ʻō-*. This term also means ‘center, setting (as of a ring),’ and *pona* also means ‘socket,’ among other meanings.

111. The Eyebrow

Representation: 92%
Motivated: 44.6%
Thereof Analyzable: 35.3% Thereof Colexifying: 9.3%
Thereof by Contiguity: 15.7% Thereof by Similarity: 9.8%
Recurrent associated meanings: eye, hair, eyelash, fur, feather, forehead, edge/fringe, wool, bone, brow of hill, eyeridge, joint

Motivated terms for ‘eyebrow’ are overwhelmingly of the lexical type, with ‘eye’ typically being one of the constituents. Unsurprisingly, the meaning of the other consituent is often ‘(body) hair,’ as in Highland Chontal *libimi gahu* /libimi lahu/ ‘body.hair eye.’ Such terms are also found in Kanuri, Koyraboro Senni, Ngambay, Berik, Dadibi, Kwoma, Toaripi, Badaga, Sora, Chickasaw, San Mateo del Mar Huave, Itzaj, Yuki, Cayapa, Jarawara, Miskito, Piro, Ancash Quechua, Rama, Bislama, Lenakel, White Hmong, and Yay. This pattern is also found in Tsafiki *ca'cá chidé fu* ‘eye tree/bone body.hair/fur/feather.’ In fact, an association with ‘bone’ is also attested in Ket and Pawnee (Pawnee, for instance, has *kirikiisuˀ* /kirik-kiis-uˀ/ ‘eye-bone-NOM’). The Pawnee term also denotes the ‘superciliary ridge, supraorbital bone,’ and this may be the explanation for the association in the other languages as well (for instance, *ca'cá chidé* may be the term for the ‘ciliary’ in Tsafiki). Similarly, One has *namnaalo palla* ‘eyeridge body.hair’ (Buin directly colexifies ‘eyebrow’ and ‘eyeridge’ inter alia), Hawaiian has *kuʻekuʻe-maka* ‘joint/elbow/wristbone/knuckle-eye,’ and Tetun *matan-fukun* ‘eye-knot/joint/knuckle.’ There are also complex terms where the constituents are ‘eye’ and ‘fur’ (although it cannot be excluded that, by colexification of ‘fur’ and ‘body

hair,' these in fact could be subsumed under terms associated with '(body) hair'). Terms betraying an association with fur, either because this is the single meaning of the constituent as in Kolyma Yukaghir *aŋd'ən-pugulbe:* 'eye-fur,' or more frequently because 'fur' is colexified with 'body hair' are found in Buli, Badaga, Berik, Kwoma, Chickasaw, Kashaya, Rama, Tsafiki, Wayampi, Lenakel, Samoan, and Yay. In many of these languages, the relevant consituent also colexifies 'feather.' This association is present in Buli, Berik, Kwoma, Chickasaw, Highland Chontal, Rama, Tsafiki, Bislama, Lenakel, Yay, while Hani has *miav-hao* 'eye-feather.' Moreover, in Sora and Samoan, there is also an association with 'wool' due to colexifying structures in constituents of complex terms. Furthermore, Hupda has *kəwəg-pə̆w* 'eye-edge,' and, similarly, Cubeo *ẽca-me* 'fringe-CLASS.LIKE.THREAD,' while Abzakh Adyghe directly colexifies 'eyebrow' and 'edge.' Colexification of 'eyebrow' with edges of specific objects is attested for Hausa and with 'edge of mountain' (as well as 'projection in cliff') in Khoekhoe (compare the parallel association between 'eyebrow' and 'mountain' evidenced by cognates in Irish, Latvian, and Albanian, Buck 1949: 220). Other complex terms where one of the constituents is 'eye' are: Efik *nditan-ënyïn*, seemingly analyzable as 'stupidity-eye,' Anggor *himboari-siri* 'eye-root/clan/origin,' Badaga *kaṇṇu eme* 'eye lid/lash,' Basque *betgain* /begi-gain/ 'eye-above' (this may be calqued, given that French *sourcil* and Spanish *ceja* go back to Latin *super-cilium* 'above-eyelid'), Lake Miwok *ṣút límme* 'eye brush,' Lakhota *ištáȟe* /ištá-ȟé/ 'eye-mountain,' Arabela *namijia-qui* 'eye-CLASS.CLOTH,' Hupda *kəwəg b'ɔk cŭ'h* containing *kəwəg* 'eye' and *b'ɔk* 'skin,' Samoan *tuā-mata* 'beyond/across-eye,' and Takia *mala-n ddawe-n* 'eye-3SG handle-3SG.' There are semianalyzable terms in many languages: Bakueri, Koyraboro Senni, Kemtuik, Lavukaleve, Kosarek Yale, Abzakh Adyghe, Bezhta, Blackfoot, Carrier, Chickasaw, Lake Miwok, Nez Perce, Santiago Mexquititlan Otomí, Pipil (Santo Domingo de Guzmán dialect), Xicotepec de Juárez Totonac, Tuscarora, Wappo, Yana, Aymara, Bora, Macaguán, Hawaiian, Malagasy, Manange, Rotuman, and Sedang.

Several of the terms mentioned so far colexify 'eyebrow' with 'eyelash' (and there is semantic shift between these two meanings in Indo-European, Buck 1949: 219). This is the case in Buli, Koyraboro Senni, Berik, Kwoma, Kyaka, Sora, Chickasaw, Itzaj, Kashaya, Lake Miwok, Cayapa, Imbabura Quechua, Tehuelche, Bislama, Lenakel, Hani, Manange, White Hmong, and Yay. Furthermore, this pattern of colexification occurs in Burarra, Ngaanyatjarra, Nunggubuyu, Tasmanian (Middle-East and Southeast), Cheyenne, perhaps in Highland Chontal, Pipil, and Chayahuita by terms with no internal morphological structure.

However, there are also terms with constituents meaning 'hair' where the second element is not 'eye.' Thus, alongside a semianalyzable term in Cahuilla, Kyaka has *lembaki-sa emanji ~ lembaki-si yamanji* 'eyelid-LOC body.hair/fur,' San Lucas Quiaviní Zapotec *gyihch lahg ~ gyihch cyi'lahg* /gyihch cyi'lahg/ 'hair eyebrow.ridge,' Tehuelche *kaščʼex ~ kaščex*, containing *-aš* 'in, inside' and *čʼex ~ čex ~ čʼe:x* 'body hair, wool' (alongside the similar term *ʔaščʼex ~ ačʼex ~ ʔačx ~ ʔačʼex ~ ʔaščex ~ a:čʼex ~ a:čex*, which also means 'beard'), and Yanomámi *wëyomahikɨ kõi* 'superciliary.arch body.hair.' The Embera and Wayampi terms betray an association with 'forehead': *dráthu-kára* 'forehead-body.hair' and *apɨkã-l-a* 'forehead-of-hair' respectively, while 'forehead' and 'eyebrow' are colexified in Rama and

Sedang (see Buck 1949: 219 on diachronic connections in Indo-European). Complex terms where one of the constituents is 'forehead' are also found in Comanche (*ka?ibʉhʉ ~ ka?ibʉʉ*, analyzable as /ka?i-pʉhʉ/ 'forehead-fuzz') and Kiliwa (*mi?=kw-m-puul=ny+mi?* 'forehead=OBJ-??-hat=POSS+forehead') alongside semianalyzable terms in Biloxi and Carib. Imbabura Quechua has *ñawi milma* 'face wool,' and a semianalyzable term where the identifiable constituent means 'face, eye' is found in Yana, and one where it means 'face, forehead' in Upper Chehalis.

Other associations include: Rotokas *uvu keru* seems to contain *uvu* 'sense something, hear' and *keru* 'harden like bone,' while Sko *lúbi* is also used with the meaning 'temple.' Welsh *ael* also means 'aisle' and 'litter.' Cheyenne *vé'henôse* is related to a verb meaning 'be singed,' and Haida *sk'yaayi ~ sk'yah ~ sk'y@s* is also used with the meaning 'the first step down into a housepit.' Arabela *susuque*, containing *-que* 'cloth,' also means 'cloth with moths,' and the relevant Cubeo term also denotes the 'tilde' in typography. Guaraní *tyvyta* appears to contain *tyvy*, which can mean either 'younger brother' and 'grave,' or *ty* 'urine.' Fijian *vacu* is also used to refer to "the place of insertion of the legs of a crab into its shell" and means 'to punch with the fist' as a verb, Kapingamarangi *himada* contains *mada* 'to look, see,' and Vietnamese *mày* is also the personal pronoun for the second person singular. Finally, Lesser Antillean Creole French *sousi* also means "care, solicitude, anxiety."

112. The Eyelash

Representation: 87%
Motivated: 61.4%
Thereof Analyzable: 53.6% Therof Colexifying: 8.0%
Thereof by Contiguity: 28.3% Thereof by Similarity: 9.2%
Recurrent associated meanings: eye, hair, eyebrow, fur, feather, eyelid, wool, leaf, brush, beard

As with the 'eyebrow' discussed in section 111 terms of the lexical type, with constituents meaning 'eye' and '(body) hair' are frequent for the 'eyelash' as well. Terms such as Hupda *kəwəg pã´t* 'eye hair' are found in Bakueri, Efik, Koyraboro Senni, Berik, Dadibi (*gedu páde nisi* 'eye near/almost hair,' as opposed to *gedu nisi* 'eyebrow'), Kwoma, Muna, Meyah, Toaripi, Kosarek Yale, Abzakh Adyghe (with an additional element present: *nebżəc* /nebż(e)-c(e)/ 'eye-horn-hair/fur/feather/wool'), Basque, Japanese, Ket, Sora, Carrier, Chickasaw, Highland Chontal, Itzaj, Kiowa, Pawnee (*kiriktaacpickiic*, the relevant term, contains *kirik-* 'eye,' *raac-* 'pubic hair,' and *kiiʔac* 'long;' it also denotes the 'persimmon'), Santiago Mexquititlan Otomí, San Lucas Quiaviní Zapotec, Aguaruna, Bororo, Cavineña (*yatuca cuatsa tsaru*, containing *yatuca* 'eye,' *tsaru* 'hair' and presumably *e-cuatsa* 'mouth'), Cayapa, Cubeo, Embera (*dau-i-kára* 'eye-border-body.hair'), Maxakalí, Piro, Rama, Tsafiki, Wayampi, Yanomámi, Bislama, Great Andamanese, Hani, Lenakel, Malagasy, White Hmong, Sedang, Tetun (where the term is figuratively also used with the meaning "opportunity, chance"), and Yay. Very similar metaphor-driven terms are found in Lavukaleve (*lemi ohal* 'eye leaf'), Samoan (*lau-mata* 'leaf eye'), and Fijian (*bebekanimata* /bekabeka-nimata/ 'coconut.leaves-POSS-eye;' for this term, note also that 'eyelash' and "cocoa husk

[sic!]" are colexified in Rotokas). The association with 'leaf' is also present in Chickasaw, due to colexification of 'body hair,' 'fur,' and 'feather.' Note also Lake Miwok *ṣút límme* 'eye brush' and Bora *hállulli* /hálluu-lli/ 'eye-CL.brush.' Katcha has *miziri m-iye* 'eyebrow ??-eye,' Arabela *namijia susuque* 'eye eyebrow,' and Vietnamese *lông mi* 'body.hair eyebrow' (an optional complex term of this kind is also found in Cheyenne, where 'eyebrow' and 'eyelash' are colexified). Other complex terms of the lexical type where 'eye' is a constituent are Mbum *tĭ-yâr* 'coming.out.from-eye,' Anggor, curiously, *ɨmboarɨ yapɨsendɨ* 'eye lid,' Badaga *kaṇṇu hoḍe* 'eye bush/thicket/wood,' Greek *matóklado*, which is perhaps analyzable as /mát-o-klado/ 'eye-STEM.FORMATIVE-branch,' Kolyma Yukaghir *aŋd'ə-šepil'* 'eye-door' (denoting the 'upper eyelash' and 'upper eyelid' specifically), Cahuilla *puščávay* /-puš-čávay-a/ '-eye-fall-??,' Kashaya *huʔu· pitemʔ*, analyzable as /huʔuy pitemʔ/ 'eye droop.of.eyes,' Jarawara *noki masiri/noko masiri* 'eye grass/bird.species,' Miskito *won nakra taya* 'BODY.PART eye skin' (this term also denotes the 'eyelid'), and Wichí *tottefwis*, containing *tef* 'eyes' and *wis* 'larvae.' There are terms with 'fur' as the second element alongside 'eye' (which is frequently colexfied with 'hair') in Buli, Berik, Kwoma, Kyaka, Meyah, Muna, Toaripi, Kosarek Yale, which also colexifies 'beard,' Abzakh Adyghe, Chickasaw, Kashaya, Lesser Antillean Creole French, Huambisa, Lenakel, Samoan, Tetun, and Yay, while Welsh has *blew yr amrant* 'hair/fur GEN eyelid'). In Buli, 'hair' and 'feather' are colexified. This is also the case in Berik, Kwoma, Meyah, Muna, Toaripi, Kosarek Yale, Abzakh Adyghe, Chickasaw, Huambisa, Tsafiki, Bislama, Lenakel, Samoan, Sedang, Tetun, and Yay, so that the relevant terms in these languages also bear an association with 'feather;' Hani has *miav-hao* 'eye-feather' for 'eyebrow' and 'eyelash,' without colexification of 'hair and feather.' Imbabura Quechua *ñawi milma* is analyzable as 'face wool,' and due to colexification with 'hair' and/or 'fur,' the association with 'wool' is also present in Abzakh Adyghe, Basque, Sora, Huambisa, and Samoan. There are semianalyzable terms involving a constituent 'eye' in Khoekhoe, Koyraboro Senni, Kosarek Yale, Bezhta, Khalkha, Blackfoot, Upper Chehalis, Chickasaw, Ineseño Chumash, Comanche, Haida (colexifying "ray of the sun shining through clouds"), Kiliwa, Lakhota, Tuscarora, Wappo, Wintu, Yana (where 'eye' and 'face' are colexified, the reference of the term is considered dubious), Copainalá Zoque, Kaingang, Manange, and Takia.

As already seen in some of the terms just mentioned, associations between 'eyelash' and 'eyelid' are also common. They are colexified in Swahili, Gurindji, Badaga, Kolyma Yukaghir ('upper eyelash' and 'upper eyelid' specifically), Upper Chehalis, Santiago Mexquititlan Otomí, Cavineña (by analyzable terms containing 'eye' and 'hair' in both aforementioned languages), Miskito, and Hawaiian (by the term *lihilihi*, the reduplication base of which means 'edge,' colexifying also 'lace,' 'to crochet,' and denoting a variety of sweet yam). Similar to the Welsh term already mentioned but without colexification of 'hair' with 'fur,' terms containing elements meaning 'hair' and 'eyelid' are also attested in Yoruba and Kyaka. Similarly, Embera has *daúiikará* which appears to be analyzable as /dauí-ikaráa/ 'eyelid-beard,' and Guaraní has *tope-a* 'eyelid-fruit.' San Lucas Quiaviní Zapotec has *gyihch bàa'ah* 'hair eyeball' (this term also denotes a card game), Tehuelche *kašč'ex ~ kaščex*, containing *-aš* 'in, inside' and *č'ex ~ čex ~ č'e:x* 'body hair, wool' (there also is the similar term *ʔašč'ex ~ ač'ex ~ ʔačx ~ ʔač'ex ~ ʔaščex ~ a:č'ex ~ a:če* which also means

'beard'), and semianalyzable terms with 'hair' are in addition present in Aguaruna, Carib, and Lengua. Several of the terms mentioned so far colexify 'eyelash' with 'eyebrow' (see Buck 1949: 219-220 for Indo-European evidence for this association). This is the case in Buli, Koyraboro Senni, Berik, Kwoma, Kyaka, Sora, Chickasaw, Itzaj, Kashaya, Lake Miwok, Cayapa, Imbabura Quechua, Tehuelche, Bislama, Lenakel, Hani, Manange, White Hmong, and Yay. Furthermore, the association is in addition found by unanalyzable terms in Bararra, Ngaanyatjarra, Nunggubuyu, Tasmanian (Middle-eastern and Southeastern), Cheyenne, Pipil, and Chayahuita, and it may be present in Highland Chontal, where the terms differ only in one segment which may be a typographical error.

Given that there are many languages without colexification of 'eyebrow' and 'eyelash,' which, however, still express both meanings with complex terms on the basis of 'eye,' it is interesting how 'eyebrow' and 'eyelash' are differentiated. One solution, adopted in Toaripi, is to use the lexical differentiation between different types of hair present in the language: *ofae ve mehe* 'eye POSS hair' is 'eyelash,' while *ovo-tui* 'eye-hair.of.head' is 'eyebrow.' Another option, present for instance in Badaga, is to have a metaphor-driven term for one of the meanings: *kaṇṇu hoe* 'eye bush/thicket/wood' is 'eyelash,' and *kaṇṇu mailu* 'eye hair/fur' is 'eyebrow.' Yet another one, present for instance in Basque, is to have a consituent neither meaning 'hair' nor being metaphorically related to it: *betile* /begi-ile/ 'eye hair' is 'eyelash,' and *betgain* /begi-gain/ 'eye-above' is 'eyebrow.' Moreover, Samoan has *fulu-mata* 'fur/wool/feather-eye' for 'eyelash,' and *fulufulumata*, with the word for 'fur, wool, feather' reduplicated, for 'eyebrow.' Finally, terms for both meanings may feature consituents meaning 'eye' and 'hair,' but one of them, typically 'eyebrow,' has another additional consituent. Thus, San Mateo del Mar Huave has *miyeed oniiügueran* 'body.hair eye' for 'eyelash,' and *miyeed opech oniiügueran* for 'eyebrow.'

Other associations include: Mali *sachong angēt kēseng* is analyzable as 'vision its those.particular.long.ones,' while Oneida *oʔnekʌ́htalaʔ* also means 'strawberry top.' The Norton Sound dialect of Central Yup'ik has *qelemyaq ~ qelemsaq*, containing a verb root meaning 'to close eyes,' while the Arabela term *maque-teja* is analyzable as 'sleep/dream-INSTR.' Rotuman *lekleki* also denotes a kind of tree.

113. The Eyelid

Representation: 71%
Motivated: 73.7%
Thereof Analyzable: 68.9% Thereof Colexifying: 5.3%
Thereof by Contiguity: 29.4% Thereof by Similarity: 20.8%
Recurrent associated meanings: eye, skin, peel/rind/shell, bark, leather/hide, eyelash, cover/lid, husk, lip, edge, surface, hair, scale, eyebrow, on, door/entrance, roof

Most frequently, motivated terms for 'eyelid' in the languages of the sample are analyzable and of the lexical type, consisting of elements meaning 'eye' and 'skin,' as in Yir Yoront *mel-pertn* 'eye-skin.' Such terms are (sometimes with additional meanings colexified with 'skin' to be discussed below) also attested in Efik, Hausa, Mbum, Ngambay, Dongolese Nubian, Berik, Kwoma, Sahu, Basque, Ket, Sora, Biloxi, Cheyenne, Kashaya, Kiowa, Lesser

Antillean Creole French, Pawnee, Pipil (Santo Domingo de Guzmán dialect), Yuki, San Lucas Quiaviní Zapotec, Copainalá Zoque, Aguaruna, Bororo, Carib, Cayapa, Hupda, Maxakalí, Miskito, Piro, Imbabura Quechua, Rama, Tsafiki, Wichí, Great Andamanese, Hani, Malagasy, Mandarin, White Hmong, Takia, Tetun, and Yay; Cubeo has a term derived from 'eye' by means of a classifier for coverings, shells, skin, etc., which however, also has a lexical counterpart of the same phonological form. Due to colexification with 'bark' in many languages (section 135), the Efik, Kwoma, Sahu, Basque, Biloxi, Pipil, Copainalá Zoque, Bororo, Cayapa, Cubeo, Maxakalí, Piro, and Tsafiki terms for 'eyelid' also betray an association with this meaning, and due to colexification with 'leather,' this meaning is associated in Dongolese Nubian, Sora, Lesser Antillean Creole French, Pipil, Copainalá Zoque, Cavineña, Miskito (which also colexifies 'feather'), Piro, Tsafiki, Wichí, and Wichí. However, for 'bark,' there are also terms where it is genuinely 'bark' rather than 'skin' which figures as a consitiuent of the term: Abzakh Adyghe has *na-pʷʔe* 'eye-bark/shell,' and such terms, without colexification of 'skin,' are also found in San Mateo del Mar Huave and Sedang, and by direct colexification in Guaraní. Further, the Efik, Basque, Sahu, Itzaj, and Sedang terms also betray an association with 'husk,' and the Efik and Biloxi terms also with 'scale.' The association with 'rind,' 'peel,' and/or 'shell' is also present due to colexification with 'skin' and/or 'bark' in Efik, Basque, Itzaj, Lesser Antillean Creole French, Pawnee, Copainalá Zoque, Cayapa, Cubeo, Maxakalí, Piro, Tsafiki, and Hani, while in Bora, the term for 'eyelid,' *hállumɨ̈ho*, is derived from *hálluu* 'eye' by the classifier *-mɨ̈ʔo* for hard shells and Hawaiian has *kuapoʻi-maka*, with *maka* meaning 'eye' and *kuapoʻi* "shell on back of crab or turtle" as well as 'kneepan,' among other meanings.

Hawaiian also has another term for the 'eyelid,' *ʻūpoʻi maka* 'cover/lid eye,' and such terms (with 'cover' either being nominal or verbal) are also found among the languages of the sample in Dongolese Nubian, Sora, Welsh, Haida, Lesser Antillean Creole French, and Yana; the connection to a verb meaning 'to cover' is also etymologically detectable in Ineseño Chumash, and Piro colexifies 'lid' with 'surface' in general, for which compare Haida *xang 7un* /xang 7unna/ 'eye top/surface.' Chayahuita has *ya'pira yonsan* 'eye edge.' An analogous term is found in Yanomámi, and Hawaiian has *lihilihi*, reduplicated from *lihi* 'edge,' which colexifies also 'lace,' 'to crochet,' and denotes a variety of sweet yam. Kiliwa has *yuw=ha?=hiiy* 'eye=mouth=edge/border,' with *yuw=ha?* meaning 'face' and *ha?=hiiy* 'lip.' Similarly, Kosarek Yale has *heing bam* 'eye lip,' a term with such structure is also found in Wappo, while Cashinahua directly colexifies 'eyelid' with 'lip.' Ngambay has *dɔ̀-kèm* 'head/on-eye,' Nikvh *njaχ-tʻxy* 'eye-on,' Buli *num-gbong* 'eye/platform-roof' and Bezhta *häyš ƛ'äq'e* 'eye.gen roof.' Curiously, One has *namna palla* 'eye body.hair,' and such a term is also found in Santiago Mexquititlan Otomí, where it colexifies 'eyelid' with 'eyelash,' as well as in Cavineña. Colexification of 'eyelash' and 'eyelid' is furthermore found in Swahili, Gurindji, Badaga, Kolyma Yukaghir (by the term *aŋd'ə-šepil'* 'eye door,' denoting 'upper eyelash' and 'upper eyelid' specifically; Dongolese Nubian also has a term where the second constituent next to 'eye' is 'door'), Upper Chehalis, Santiago Mexquititlan Otomí, Cavineña, Miskito, and Hawaiian (by the term *lihilihi* mentioned above).

Other complex terms for the 'eyelid' on the basis of 'eye' are Efik *mfut-ënyïn* 'shade eye,' Kanuri *fártà shîm-bè* 'root eye-of,' Kildin Saami *čall'm-rūmtas* 'eye-brim,' Blackfoot *ootokiáápinihpis*, perhaps containing *ootoki'p* 'skull' and *aapini* 'eye,' Cahuilla *pušyúmuve?*, literally 'eye hat,' Upper Chehalis *ƛúǩʷ=ičn-s t mús* 'above/top=ridge=-?? INDEF.ART eye/eyeglasses' for 'upper eyelid' specifically, Kiliwa *yuw=nat-u?=kw-waa* 'eye=top/atop-OBL=WH-sit,' Chayahuita *ya'pira sha'shatë'* 'eye corner.of.mouth,' Embera *dau-í* , apparently analyzable as 'eye-wing,' Toba *l'ai lapo'te* 'upper eyelid,' containing *'ai* 'eye' and *apo'* 'poncho,' *l'ai'te lqa'* 'lower eyelid,' containing *l'ai'te* 'eye' and *lqa'* 'chin,' and Fijian *daku-daku-ni-mata* 'back-RED-POSS-eye' for the 'upper eyelid' and *dreke-ni-mata* 'hollow/cavity-POSS-eye' for the 'eye socket' and 'lower eyelid,' Lenakel *nouanhal-nimr-* 'egg-eye-,' and Manange *2mi-2pʰi* 'eye-up.' Finally, Tetun has *matan-kukun* 'eye-nail/claw,' and due to colexification with 'skin,' the association with 'fingernail' is also found in Biloxi. Semianalyzable terms where the identifiable constituent is 'eye' are found in Bakueri, Yoruba, Toaripi, Waris, Sora (for the 'lower eyelid' specifically), Carrier, Upper Chehalis, Highland Chontal, Comanche, Santiago Mexquititlan Otomí, Xicotepec de Juárez Totonac, Wappo, Cavineña, Hani, and Rotuman.

Furthermore, the Oneida term *okahkwilo?ókta?* likely revolves around the verb root *-kahkwilo?ok-*'to blink,' and similarly, Khalkha *anisxa* may be related to the verb *aniski-* 'to blink eyes repeatedly.'

Other associations include: Muna *bhangkoa* also means 'eye socket' and, verbally, "to roll up something broad but thin." *Zobki*, a Khalkha term for 'eyelid,' also means 'corner of the eye' (compare *zobkis* 'corner'). Ineseño Chumash *'ixma'y* also means "to have a film or cloud in the eye, cataract" and Guaraní colexifies 'eyelid' with 'petal.'

114. The Finger

Representation: 96%
Motivated: 59.6%
Thereof Analyzable: 22.1% Thereof Colexifying: 37.5%
Thereof by Contiguity: 20.5% Thereof by Similarity: 34.7%
Recurrent associated meanings: hand, toe, arm, claw/paw/forefoot, child/son, branch out, little, fingernail, branch, fruit, fork, hoof, head, neck

'Finger' and 'hand' are often not distinguished lexically (Brown 2005a), and this is also true of many sampled languages, namely Ngambay, Burarra, Kyaka, Ngaanyatjarra, Nunggubuyu, Tasmanian (Western, Middle-Eastern and Southeastern), Yir Yoront, Abzakh Adyghe, Cahuilla, Upper Chehalis, Cheyenne, Kashaya, Kiliwa, Nez Perce (where the narrow interpretation 'finger' is restricted to compounds), Oneida, Wintu, Yana, Arabela, Cavineña, Chayahuita, Jarawara, Great Andamanese, and Hawaiian. In a subset of these languages, the same term is also used for the 'arm' as a whole. This is the case in Ngambay, Kyaka, Ngaanyatjarra, Cahuilla, Kiliwa, and Hawaiian, where the relevant term also means 'sleeve' and 'five.' However, it is very often also the case that there are complex terms for 'finger,' often metaphor-driven, where one of the constituents is 'hand,' which is sometimes colexified with 'arm' generally (Brown 2005b). None of these metaphor-driven pat-

terns is overwhelmingly frequent, but they do recur on a global scale with some strength. The most common meaning of the second constituent in complex terms is 'child,' as in Katcha *bibala ma nizo* 'child GEN hand.' Such terms are also found in Koyraboro Senni, Mbum, Gurindji, Chickasaw, Itzaj, Pipil, and Maxakalí (in Chickasaw and Pipil with 'son' specifically rather than 'child' generally). Perhaps relatedly, in Yana, *tidʔyau(na)* is also a kinship term for 'man's younger brother' and 'woman's younger sister.' A word meaning 'small' or 'little' also figures in some languages, for instance in Miskito *mita sirpi* 'hand little.' Similar terms are found in Carib, Piro, and Samoan. In Mali, the word for 'finger' is derived from that for 'hand' by means of a masculine noun class prefix, which is associated with smallness (Stebbins 2005), and Yir Yoront, which colexifies 'finger' and 'hand,' has the optional complex term *yor-mart* 'hand/finger-little/tiny' to single out the 'finger'-reading. In the Piro term, the second constituent conveying 'smallness' has also a reading 'fruit' (the association between diminutive semantics and 'fruit' is also reported by Matisoff 1992). Similarly, Lenakel has *noua-nelm-* 'fruit-hand/arm-,' and Tetun *liman-fuan* 'arm/hand-fruit/heart.' In the Carib term, the constituent *sikɨlɨ* is glossed more precisely as 'little end,' and similarly, Kapingamarangi has *madaalima* /mada-lima/ 'end-hand/arm.' Hawaiian *manamana*, colexifying 'finger' with 'claws, rays, forks' inter alia, is reduplicated from *mana* 'branch,' Malagasy has *rantsantànana*, analyzable as /ràntsana-tànana/ 'branch-hand,' and 'branch' and 'finger' are colexified directly in Aguaruna. Relatedly, Nuuchahnulth has *ċaċałaqɬukum* /ċa-ċałaq-nukʷ-im/ 'RED-branch.out-in.or.at.the.hand,' and similar terms highlighting that the fingers protrude from the hand are featured in Kaluli and Yay (*dagi gasa* 'hand/paw things.that.separate' and *ɲiaŋ*[5] *fuŋ*[4] 'to.come.apart hand' respectively) alongside a reduplicated term with the meaning 'branch, branch out' in Hawaiian. Similarly, Tehuelche and Hawaiian colexify 'finger' with 'fork.' Pawnee has *iks-kiic-uʔ* 'hand-neck-NOM,' and Bororo *era-ko* 'hand-neck.' Other complex terms where one of the constituents is 'hand' and/or 'arm' are Noni *kpwan ɛbo* 'to.mark hand,' Baruya *a'jawɨnya* /ata-jawɨnya/ 'hand-fish' (also denoting the 'thumb' specifically; note that in other languages such as Khoekhoe, 'thumb' is explicitly not included in the extensional range of the 'finger'-word), Muna has *wuna-no lima* 'flower-POSS hand/arm,' Sko *nò-kang-kang* 'arm/hand-tusk/eat-RED,' Abzakh Adyghe *ʔe-pe* 'hand-tip' (compare possible connections to this meaning in Ancient Greek as well as cognates between Baltic and Germanic suggestive of the association, Buck 1949: 240), Bora *méhójtsɨwa* /mé-hójtsɨ̃-gwa/ '1PL-hand-SCM.2d.straight,' Cubeo *ãmu-yo* 'hand/arm-CLASS.LARGE.CYLINDRICAL.SLIM.AND ACUTE.OBJECT,' Cayapa *tyamishu* /tyaapa-mishu/ 'hand-head,' a term with identical structure in Tsafiki, Kaingang *nĩgé féj* 'hand leaf' and *nĩgé juféj* 'hand long.things,' and Bwe Karen *-cu-yò* '-arm/hand-point.at' (note also that Rotokas *piiroo* also means "point towards something with finger or object;" this association is probably also at the heart of Latin *digitus*, Buck 1949: 240).[9] Furthermore, Yanomámi *imi-hena* is analyzable as 'carry.in.hand-leaf.' Semianalyzable terms are featured in Katcha, Dadibi, One, Sentani, Abzakh Adyghe, Biloxi, Carrier, San Mateo del Mar Huave, and Wappo.

[9] To this list of metaphorical denominations one can add, from outside the sample, Dene Sųłiné's *denelatthałé* 'person's hand-awl' for 'finger' and *denelachédh* 'person's hand-duck' for thumb (Rice to appear).

Furthermore, Ngaanyatjarra, Yir Yoront, Basque, Upper Chehalis, Cheyenne, Lengua, Rama, Hawaiian, and Rotuman colexify 'finger' with the homologous structures in animals, namely 'claw,' 'paw,' or 'forefoot.' Similarly, in the Nunivak Island dialect of Central Yup'ik, *ipik ~ ipi* may also be used with reference to the "limb of quadruped or insect." Kwoma colexifies 'finger' with 'person' (as well as, by extension, 'twenty' and other meanings), and the Hooper Bay and Chevik dialect of Central Yup'ik has *cugaraq ~ cuaraq ~ yuaraq* containing *cuk ~ yuk* 'person.' Basque *behatz*, containing *hatz* 'finger,' also means 'hoof,' 'vestige, trace, mark' and other things, ('hoof' is also colexified in Lengua), while the Buli and San Lucas Quiaviní Zapotec terms are also traditional units of measurement. Hausa, Dadibi, and Basque colexify 'finger' with 'fingernail' (Hausa also with "[d]ermatoid cyst under upper eyelid of horse").

In addition, when simplex terms for 'finger' do exist (sometimes exhibiting other patterns of colexification as well), they can often at the same time be used with reference to the 'toe' (also in some Indo-European languages, Buck 1949: 240-241), with disambiguating complex terms in some languages. This is true of Buli, Efik, Rendille, Swahili, Yoruba, Kwoma, Lavukaleve, Rotokas, Northeastern Tasmanian, Toaripi, Sahu, Waris, Badaga, Bezhta, Japanese, Ket, Khalkha, Welsh, Blackfoot, Cheyenne, Ineseño Chumash, Haida, Tuscarora, Central Yup'ik, San Lucas Quiaviní Zapotec, Copainalá Zoque, Aymara, Cayapa, Huambisa, Lengua, Macaguán, Fijian, Hawaiian, Lenakel, Rotuman, Sedang, Takia, and White Hmong. An interesting generalization emerges, namely, that terms betraying an association between 'finger' and 'hand' and those colexifying 'finger' and 'toe' are mutually exclusive in the languages of the sample. That is, there is no language in the sample in which all three meanings are colexified (although one language may feature several terms for 'finger,' one of which features one association and another one the other). This thus appears to be a strongly dispreferred pattern cross-linguistically.

Other associations include: Berik *bola* colexifies 'forefinger' with 'ball,' Kyaka *kingi* also may refer to "name, title, rank, reputation," and Ngaanyatjarra *mara* also means 'five.' Chukchi *rəlɣəlɣən* contains *ɣəlɣən* 'skin,' Welsh *bys* colexifies 'hand of clock,' and Cheyenne *mo'ëško* 'ring.' Tuscarora *uhsú?kweh* may also be used to refer to a 'rake,' a 'thimble' or a 'herring tooth.' The bound term *wa-* in Yana is also a general prefix for long objects, while Jarawara *yehe/yehe* may also refer to the 'self.' Rama *isúluk up* contains *up* 'eye, something round' and presumably *uk* meaning 'skin, coat, rind' inter alia. *Ehethakɨ*, a Yanomámi term for 'finger,' consists of *ehetha* 'wrist' and the quantal classifier *kɨ* (for which see § 4.4.1). Rotuman *käkä'e* also means 'wall of house,' and Lesser Antillean Creole French *dwet* also means 'upright, straight, direct' and, by extension, 'righteous.'

115. The Fingernail

Representation: 95%
Motivated: 69.2%
Thereof Analyzable: 12.8% Thereof Colexifying: 56.1%
Thereof by Contiguity: 1.3% Thereof by Similarity: 46.4%
Recurrent associated meanings: claw/talon, toenail, hoof, finger, hand, shell, arm, skin, paw, bark, horn

'Fingernail' is very frequently colexified with 'claw' and/or 'talon' (see Buck 1949: 245 on Polish). This is the case in as many as 67 sampled languages, namely Bakueri, Buli, Efik, Hausa, Khoekhoe, Noni, Dongolese Nubian, Rendille, Swahili, Anggor, Buin, Burarra, Gurindji, Mali, Ngaanyatjarra, Toaripi, Waris, Yir Yoront, Badaga, Basque, Bezhta, Greek, Japanese, Ket, Khalkha, Laz, Kildin Saami, Welsh, Kolyma Yukaghir, Acoma, Upper Chehalis, Cheyenne, Highland Chontal, Ineseño Chumash, Haida, Itzaj, Lake Miwok, Lakhota, Lesser Antillean Creole French, Kiowa, Oneida, Pipil, Xicotepec de Juárez Totonac, Tuscarora, Wintu, Yaqui, Yuki, Central Yup'ik, San Lucas Quiaviní Zapotec, Copainalá Zoque, Aguaruna, Aymara, Carib, Cashinahua, Cayapa, Cubeo, Guaraní, Huambisa, Miskito, Piro, Ancash Quechua, Wayampi, Fijian, Hawaiian, Malagasy, Rotuman, and Tetun (in Rotuman, by a semianalyzable term containing 'hand;' there are other such terms in Upper Chehalis and Hani). Complex terms betraying this association are found in Wappo *meʔčúš*, containing *meʔ* 'hand, by hand' and *čúš*, also occurring in *pheʔčúš* 'claw, hoof, toenail,' and in Great Andamanese *ôngkōro-bôdoh* 'finger/hand-claw.' There is a semianalyzable term in Baruya. In a subset of these languages, Efik, Dongolese Nubian, Toaripi, Greek, Kildin Saami, Biloxi, Cheyenne, San Lucas Quiaviní Zapotec, Aymara, Arabela, Guaraní, Miskito, Ancash Quechua, Fijian, Hawaiian, Rotuman and Tetun, 'hoof' is in addition colexified; this is also the case in Muna and Basque. Furthermore, Buli, Guaraní, and Miskito colexify 'paw' and 'fingernail' in addition.

Biloxi and Rama colexify 'fingernail' with 'bark,' and a complex term is found in Jarawara (*yehe atari* 'finger/hand skin/bark/peel'). Biloxi also colexifies 'fingernail' with 'horn' (also 'scales of fish'), and Abzakh Adyghe has *ʔebżane* /ʔe-bż(e)-ne/ 'hand-horn-part/unit' (this association is also found Avestan and Ancient Greek, Buck 1949: 245). Unsurprisingly, the complex terms so far mentioned are not the only ones where one of the constituents is 'hand' or 'finger' (or 'arm'). One has *meu alo* 'arm egg.shell,' Toaripi *mai haro* 'hand/arm head/shell,' Sko *nòbi* /nò-bí/ 'arm/hand-shell,' Hupda *cob b'ɔ́k* 'finger-skin/shell,' Bislama *sel fingga* 'shell finger,' Samoan *atagi lima* 'shell arm,' and Yir Yoront colexifies 'fingernail' with various kinds of shell directly (note also that Hausa *k'umba* also denotes the 'shell of a mussel' inter alia). For the association with 'skin,' note also that Biloxi and Rama colexify 'fingernail' and 'skin,' as well as the complex terms *tofwefwt'oj*, containing *fwefw* 'finger' and *t'oj* 'skin, hide' in Wichí and *imi-si* 'carry.in.hand-cover/skin' in Yanomámi. There is a semianalyzable term involving a constituent 'skin' in Chukchi, and a semianalyzable one with a constituent meaning 'skin, bark, shell' in Guaraní. Other complex terms are Kwoma *tapa bi* 'arm/hand point,' Blackfoot *awó'taanookitsis* /awo'taan-mookitsis/ 'shield-finger/toe,' Tsafiki *tehué* /tede-hué/ 'hand-fast,' Manange *1ja-2ʃiŋ* 'hand-wood,' and Mandarin *zhi3-jia1* 'finger-armor.' Hausa, Dadibi, and Basque directly colexify 'finger' and 'fingernail,' and semianalyzable terms involving either 'finger,' 'hand,' and/or 'arm' are furthermore attested in Mbum, Dadibi, Kwoma, Carrier, Chickasaw, Copainalá Zoque, Cayapa, Maxakalí, and Rotuman.

Finally, as is the case with 'finger' and 'toe' (compare section 114), 'fingernail' and 'toenail' are often colexified, at times with the possibility of disambiguating complex terms. Terms explicitly glossed as also capable of referring to the 'toenail' are found in Buli, Efik, Khoekhoe, Rendille, Anggor, Buin, Gurindji, Kaluli, Kyaka, Lavukaleve,

Ngaanyatjarra, Nunggubuyu, Rotokas, Tasmanian (Northeastern, Middle-Eastern, and Southeastern), Sahu, Waris, Basque, Chukchi, Ket, Kildin Saami, Sora, Biloxi, Blackfoot, Upper Chehalis, Cheyenne, Highland Chontal, Itzaj, Oneida, Wintu, Central Yup'ik, Cayapa, Embera, Piro, Bislama, Fijian, Hawaiian, Lenakel, Rotuman, and Tetun.

Other associations include: Nunggubuyu *yara* also means 'to smell, detect, sense something,' while Muna *konisi* also denotes a kind of crab. Ngaanyatjarra *miltji* also denotes the "spines of thorny devil" and means 'rake.' The Kosarek Yale term *aluk* also is used as the name of a variety of banana, and Yir Yoront *yorwel* also means 'spoon.' Greek *níchi* is also used figuratively with the meaning 'clutches,' and Ket colexifies 'fingernail' also with 'needle, pin.' Cahuilla *sálul* is presumably related to the verb *-sáluk-* 'to scratch,' the Nuuchahnulth term *čałča* contains the verb *čał-*, meaning 'to get split lengthwise,' while Wintu *ḵah* also means 'wind, blow.' Wayampi *pãpɛ̃* contains *ãpɛ̃* 'to bend' and also denotes 'vaned feathers,' Fijian *kuku* is also the name for a kind of mussel, while Bwe Karen *θɔ́mi* is also the name of a shell fish species. The Kapingamarangi term *madaniha* consists of *niha* 'tooth' and the augmentative prefix *mada-*, and Tetun *kukun* colexifies 'fingernail' with 'pincer of crab.'

116. The Guts

Representation: 95%
Motivated: 47.7%
Thereof Analyzable: 13.0% Thereof Colexifying: 34.9%
Thereof by Contiguity: 3.9% Thereof by Similarity: 5.9%
Recurrent associated meanings: innards/bowels/entrails, belly/stomach, faeces, inside(s), rope/string/line, sausage, heart, womb, catgut, child/son, inner tube of tire, end

Many sampled languages have terms referring to the 'innards' or 'entrails' or 'bowel(s)' generally, including the 'guts' or 'small/large intestine.' These are Buli, Efik, Khoekhoe, Mbum, Ngambay, Dongolese Nubian, Yoruba, Buin (the term includes also the 'bladder' in its semantic range), Kyaka, Lavukaleve, Muna, Nunggubuyu, Sentani, Kosarek Yale, Yir Yoront, Badaga (where the relevant term is also capable of referring to the 'liver' specifically), Greek, Khalkha, Welsh, Ineseño Chumash, Haida, Lakhota, Lesser Antillean Creole French, Nez Perce, Pawnee, Tuscarora, Central Yup'ik, Abipón, Aguaruna, Arabela, Cashinahua, Embera, Toba, Fijian, Bislama, Bwe Karen, Hawaiian, Rotuman, and Samoan; in the latter two languages, the term is also a jocular designation for a 'child' or 'son' specifically. 'Guts' and 'inside(s)' are colexified in Kaluli, Kwoma, Abzakh Adyghe, Yuki, Jarawara, Lengua, Lenakel, and Sedang, while Embera colexifies it with 'innermost,' and Japanese has *nai-zō* 'inner-organ.'

As for other associations with body-parts, Badaga, Abipón, and Hawaiian colexify 'guts' with 'heart' (among other meanings in some languages, see section 117); 'heart' is in quotation marks in the Badaga source, suggesting a figurative rather than literal usage, and the same may be true of Hawaiian. Sahu, Badaga, and Sedang colexify 'guts' with 'womb' (the latter language also with "inner edge of bamboo or of kơmea square strip of rattan").

In some sampled languages, there is a metaphorical transfer realized by analyzable terms of the lexical type from 'line' or 'rope' to 'guts.' For instance, Hupda has *tok tít* 'stomach string/vine,' and in Toaripi, 'rope' and 'intestines' are colexified. Also attested in this class of terms is to have 'faeces' rather than 'belly' or 'stomach' as the second element, as in Cayapa *pe-shilli* 'excrement-line.' Such terms are also attested in San Mateo del Mar Huave and Tsafiki. Nuuchahnulth *ċiyup* contains *ċis* 'in line, rope,' and a semianalyzable term involving a constituent meaning 'thread, string' is found in Highland Chontal.

Khoekhoe, Mbum, Kwoma, Ngaanyatjarra, Southeastern and Western Tasmanian, Badaga, Basque, Khalkha, Cheyenne, Kiliwa, Lake Miwok, Guaraní, and Miskito directly colexify '(part of the) guts' with 'stomach' and/or 'belly' (Badaga also with "something that came from the stomach"). Burarra more specifically colexifies 'stomach' with 'part of intestines.' Katcha has *ɛm(a) aidhi* /ɛma aidhi/ 'object stomach,' Meyah *otkonú efesá* 'stomach child,' Carib *uwempo sikilipo* containing *uwempo* 'belly' and *sikili* 'little parts,' and Toba *lailshic*, presumably analyzable as /l-auel-shi-c/ '3SG.POSS-stomach-CLASS.PLACE.OF.PROVENIENCE-MASC' and *l-auel l-'ec* '3SG.POSS-stomach 3SG.POSS-content.' Furthermore, in Swahili, the term for 'guts' consists of that for 'stomach' and a noun class prefix, in Embera, the relevant meanings are expressed by the same term with alternating genders, and semianalyzable terms where the identifiable constituent means 'belly' and/or 'stomach' exist in Sko and Bora. Ngambay, Baruya, Tasmanian (all varieties), Haida and Jarawara colexify 'guts' with 'faeces' directly (Baruya also with 'flatulence'), Kosarek Yale has *disidik*, where *di* is 'faeces' and *sidik* can mean 'rest, leftover,' 'genuine, true,' as well as 'to straighten, sprout, grow quickly,' Bororo has *pe-guru* 'excrement-liquid,' Cubeo *cu̶ra-me* 'faeces-CLASS.LIKE.THREAD,' Piro *hitš̲ka-pi* 'excrement-rod,' and in Yanomámi, *shi-ki̶* ~ *shi-pë* consists of *shi* 'faeces' and a quantal classifier (see § 4.4.1.1.). Semianalyzable terms where the identifiable constituent means 'faeces' are attested in Berik, Kashaya, Pipil, Sáliba, and Lenakel, and ones where it means 'to defecate' in Abzakh Adyghe and Chayahuita.

Abzakh Adyghe also has the term *č'etʼəy* /č'e-tʼəy(e)/ 'end-loop,' and Kolyma Yukaghir *jömgid-i:čə* 'turn-end.' Presumably by provenience contiguity, Bezhta, Nez Perce, Tuscarora, Manange, and Rotuman colexify 'guts' with 'sausage' (the association was present in Latin by formation of a diminutive, Buck 1949: 1086), and, by metaphor, Buli and Hausa colexify 'guts' with 'inner tube of tire' (Hausa also with 'works,' e.g. as of a clock), and Basque and Lesser Antillean Creole French with 'catgut' (Basque also with a 'medicinal probe').

Other associations include: Khoekhoe *ǃnāb* also means 'interior' and 'diarrhoea.' Yoruba *oriẹ̀kọ́* contains *ẹ̀kọ́* 'head,' Dadibi *haliga* also means 'umbilical cord,' and Kwoma *sugu* is also the name of a 'digging stick.' Muna *ghule* also means 'snake, worm' and *ngallungallu* also denotes a "rattan nose-ring for cows or buffaloes." Ngaanyatjarra *tjuni* colexifies 'guts' and 'stomach' with 'front of person,' the "hollow of anything concave," the 'round part of fruit,' and is conceived of as the seat of emotions (which is also the case in Hawaiian), while Nunggubuyu *-lhanᵍaj* also denotes a "minnow, small fish (used as bait)." Another Nunggubuyu term, *–nᵍu-*, also means 'to eat,' while Waris *ungul* 'large intestine' appears to contain *ung* 'anus.' Basque *golko* also means 'bosom, breast' as well as 'bay,

gulf,' Japanese *chō* also means 'butterfly' (it is in both readings borrowed from Chinese, but goes back to different etyma), and Khalkha *gedesy(n)* is also the name of the "loop for o/u in Khalkha script;" this term may be derived from *gede* 'nape of neck, occiput.' Sora *on'loida:n ~ onlaida:n* contains *lˀoːj-* 'testicle.' Welsh *perfedd* also means 'middle,' and *pot* also 'navel.' Central Yup'ik *cakunglluut* contains the postbase (see § 4.4.2) *-nglluk* 'not too good,' and *imanaq imaq* 'contents, bullet, pus, ocean' and perhaps the postbase *–naq* 'one like.' Arabela *mana-ca* is analyzable as 'mud-CLASS.FRUIT,' Jarawara *tori/toro* also means 'streambed,' Wayampi *ɨkɛ* also 'side' and 'older sister,' Fijian *wāwā* also 'tired out,' and Bwe Karen *-bwì* also for beverages "to be potent, strong."

117. The Heart

Representation: 97%
Motivated: 48.0%
Thereof Analyzable: 11.5% Thereof Colexifying: 37.3%
Thereof by Contiguity: 4.7% Thereof by Similarity: 22.6%
Recurrent associated meanings: feel/think, center/middle/core, soul/spirit, belly/stomach, inside/interior, heart in cards, liver, courage/boldness, life/alive, breast, seed/pit, pith, lungs, breath, want/desire, kidney, love, conscience, fruit, guts, womb, wind

Terms for 'heart' are sometimes lexically connected to other internal organs of the body on the one hand, and are, on the other hand, associated with emotional states and cognitive abilities more generally in many languages.

By colexification, there are three languages, Ngambay, Guaraní, and Yanomámi, where 'heart' can also refer to the 'kidney' (and in Yanomámi to a spherical object in general), and in four, Buin, Burarra (by the analyzable term *mun-molma* 'CLASS.DOMESTIC-warmth'), Kwoma, and Yay, 'heart' and 'lungs' are colexified respectively (and there are other languages in which the latter meanings are expressed by morphologically complex terms on the basis of 'heart,' see section 122). Colexification with 'liver' is a little more frequent, occurring in Ngambay, Kwoma, Sahu, and San Lucas Quiaviní Zapotec, while in Yay, *tuaŋ²* denotes the internal organs 'heart' and 'lungs' and sometimes includes the 'liver' (note also that terms similar in shape to Nunggubuyu *andhiṟi* 'heart' are found in neighboring languages with the meaning 'liver'). Moreover, in Khoekhoe there is a very general term for 'offals,' including 'lungs, liver, kidney' and 'heart' that can also refer to the 'belly' or 'stomach.'

In addition, there is one language, Mbum, where 'heart' is *làù wârké* 'liver male.' More frequent is an association with the 'stomach' cross-linguistically. Ngambay, Abzakh Adyghe, Badaga, Laz, San Mateo del Mar Huave, Itzaj (by the term *pusik'al*, analyzable as /puus-ik'-al/ 'dusting-wind-COLL;' for this, note also Cubeo *ũme-dʉ* 'wind-CLASS.ROUNDISH.THREEDIMENSIONAL.OBJECT' and that the association is present in Tariana according to Aikhenvald 2003: 128), Santiago Mexquititlan Otomí, Yuki, Arabela, and Guaraní colexify the meanings, in which case the respective terms appear to be more vague in their reference, referring to the internal part of the trunk more generally. Similarly, Piro

colexifies 'pit of the stomach' more specifically, and furthermore, a semianalyzable term featuring a constituent meaning 'stomach, belly' as well as 'front of body' exists in Yir Yoront. Badaga, Abipón, and Hawaiian colexify 'guts' with 'heart' (Badaga also with 'dysentery,' 'womb,' and other meanings; note though that 'heart' is in quotation marks in the Badaga source, which suggests a figurative rather than literal usage, and the same may be true of Hawaiian), Katcha, Highland Chontal, Kaingang, and Great Andamanese with 'breast,' and Badaga and Welsh with 'womb,' although this reading is obsolete in Welsh. Moreover, Kwoma colexifies 'heart' with the "upper half of torso," and Sahu with 'chest.' For the associations with 'wind' in Itzaj and Cubeo just mentioned, note also that in some languages there is a connection with 'breath' or 'breathe:' Kolyma Yukaghir, Wintu, and Rotuman colexify these meanings (Wintu also 'to rest, be restored, revive'), and Rotokas has *vovou isi* 'breathe/want CLASS.ROUND.OBJECT.'

On the other hand, the heart is often conceived of as the seat of emotions or the soul. Yoruba, Lavukaleve, Sahu, Badaga, Kolyma Yukaghir, Yaqui, San Lucas Quiaviní Zapotec, Guaraní, Hupda, Rama, Tsafiki, and Bwe Karen colexify 'heart' with 'soul' and/or 'spirit' (Yaqui also with 'ghost'), and in as many as 33 sampled languages, 'heart' is lexically associated with cognitive abilities, such as 'feeling,' 'thinking,' or is considered the seat of emotions more generally. This is the case by colexification in Noni, Buin, Burarra, Kyaka, Lavukaleve, Muna, Sahu (where the relevant term also means "to tell a story"), Waris, Badaga, Khalkha, Haida, Nuuchahnulth (where the relevant term also means 'brain' and 'spinal cord'), Xicotepec de Juárez Totonac, Guaraní, Tehuelche, Tsafiki, Hani, Bwe Karen, Hawaiian, Lenakel, Mandarin, Rotuman, and Sedang, and by complex terms in Buli (*sukiri* ~ *sikiri* /sui-kiri/ 'mind-root/reason'), Upper Chehalis (*s-qwə́lm* 'CONTINUATIVE-think'), and Abipón (*n-eo-et-Ra-nat-Ra* 'POSS.INDEF/3SG-feel-COGNITIVE.ACTIVITY-ABSTR-??-ABSTR'). There is a semianalyzable term where the identifiable constituent also means 'middle' in Buin. In four languages of the Old World, Efik, Badaga, Khalkha, Welsh, and also in Lesser Antillean Creole French, associations with 'courage,' 'boldness' and other meanings are found (see Buck 1949: 251 for Indo-European), and in three languages, Waris, Bora, and Ancash Quechua, the heart is lexically associated with 'love.' In Basque, it is also a term of endearment (glossed as 'darling'). Alongside Rotokas, Khoekhoe has *ǂgao-b* ~ *ǂgao-s* 'want/desire-3SG.MASC' ~ 'want/desire-3SG.FEM,' and such associations are found by colexification in Muna, Bwe Karen, and Rotuman, and a semianalyzable term with an element meaning 'to breathe' as well as 'to want, desire' exists in Rotokas. Note also that in Kiowa, where 'heart' is *t'ę̨n*, there is also a verbal prefix *t'ę̨n-* "referring to desire." Moreover, Yoruba, Badaga, and Guaraní colexify 'heart' also with 'conscience.' Other associations with cognitive and emotional states are: Buli *kpa-ziim* 'occiput-blood' means 'heartbeat' and 'anxiety, fear' and in restricted contexts also 'heart,' Hausa *zuciya* also means "to get into a temper," "quick temper," wheras Ngambay *wùr*, in contrast, also means 'patience.'

The heart is also at times (in Sora, Nuuchahnulth, Tuscarora, Central Yup'ik, where the relevant term however is especially used for the heart of fish, and Bwe Karen) lexically connected with words for 'life' (in Sora also with 'intimacy' and 'friendship'). This is particularly frequent in North America, where the connection is typically realized by morphologically complex terms (Nuuchahnulth, Tuscarora, and Central Yup'ik). For in-

stance, a dialectal Central Yup'ik word for the 'heart' is *unguvan* /unguva-(u)n/ 'be.alive-device.for.' The association occurs by colexification in Sora and Bwe Karen.

The 'heart' is also sometimes (namely in Khoekhoe, Yir Yoront, Basque, Kiliwa, Pawnee, and Bislama) extended to mean 'heart in cards,' likely under European influence. Frequently, the heart is also extended to mean 'center, middle' or 'core' (see Buck 1949: 251 on Indo-European and Heine and Kuteva 2002: 171 for a perspective from grammaticalization). This is the case in Abzakh Adyghe, Basque, Welsh (where the same term also means 'navel'), Kolyma Yukaghir, Ineseño Chumash, Itzaj, Lesser Antillean Creole French, Tuscarora, San Lucas Quiaviní Zapotec, Arabela, Bora, Lengua, Ancash Quechua, Mandarin, and Samoan, while in Buin, there is a semianalyzable term for 'heart' where the identifiable constituent is 'middle' (or 'to know, think of, reflect on'). Similarly, 'heart' is colexified with 'inside' or 'interior' in Ngambay, Northeastern Tasmanian, Basque, Maxakalí, Ancash Quechua, Malagasy and Rotuman, while the San Mateo del Mar Huave term mentioned above is analyzable as *ɔmeaats-aran* 'inside-INAL.POSS.' Shape-based similarity, as well as similarity in the position of the two with respect to the larger structure they occur in is also responsible for associations between the heart and the meanings 'seed' or 'pit' in Sentani, Kosarek Yale, Nez Perce, and Samoan; note also that Hupda *hã́wɨg* is probably from **haŋ-wɨg* 'breath-seed.' However, 'fruit' itself is also colexified with 'heart' in Muna and Tetun (in Muna also with 'banyan tree'), and in Kosarek Yale in fact all three meanings can be expressed by a single term (as well as 'flower-stalk,' 'egg,' and figuratively 'child'). Further, Dadibi has a complex term for 'heart' involving *du* 'hill, mount, fruit.' In Ineseño Chumash, Nez Perce, Nuuchahnulth and Fijian, respective terms also can be used to refer to the 'pith' or 'center' of trees (the relevant Fijian term also denotes the breadfruit tree and its fruit).

Other associations include: the relevant Dongolese Nubian term is also used as a reflexive pronoun. Lavukaleve *vuvul* is also used to refer to a 'young coconut,' and Kosarek Yale *wangka* also in dancing songs to the "seed of sweet potato" specifically as well as, without restrictions as to register, to the "cocoon of a spider species." Abzakh Adyghe *gᵒə* is extended to 'surrounding,' and, from there on, to 'surface, territory.' Kolyma Yukaghir *šubeďə* colexifies "middle part of a fish trap, middle part of a boot" (this term is connected to a root meaning 'to run'). Cheyenne *hësta* is also the name of the 'heart,' a constellation of stars. Haida obsoletely colexifies 'heart' with 'throat,' while the relevant Wappo term contains an element meaning 'stone.' Aguaruna *anentái* is derived from *ánen ~ áneg* 'love song' by means of the instrument nominalizer *-tai*, and Bororo *bapo ~ wapo* also denotes a 'rattle.' The Chayahuita term *nino-pi* is analyzable as 'beehive-CLASS.FRUIT.BODY.OR.BACK,' while Jarawara *makawari* appears to contain *maka* 'snake, jungle monster' and *wari* 'tree.' Ancash Quechua *puywan ~ puywaq* is derived from *puyway* 'for the heart to beat.' Imbabura Quechua *tulpa rumi* is analyzable as 'hearth stone,' whereas the Sáliba term *omaĩdi* also means 'animal.' Wayampi *tulu-ãkã* appears to be analyzable as 'big-head,' Great Andamanese *ôtkûktâbana* contains *bana* 'globular,' Bislama colexifies 'hat, helmet' (due to phonological collapse of English *heart* and *hat*) and Hawaiian *puʻu-wai* is literally 'protuberance-water' and colexifies "a heart-shaped locket as of gold ... or silver," "small suckers on a taro plant," and has other figurative usages. Hani *neesiq* colexifies 'heart' with 'soy bean'

(*siq* is a classifier for round objects), Kapingamarangi *manawa* also carries the meanings 'predisposition' as well as to "come up from the surface (from diving)," Mandarin *xin1* (reflecting early middle Chinese *sim* and *sin* respectively) also means 'new,' and Sedang *ihiam* also denotes the "inside works of a machine" inter alia, and *nuih* also the "pole in roof of house."

118. The Jaw

Reperesentation in Database: 79%
Motivated: 57.6%
Thereof Analyzable: 20.4% Thereof Colexifying: 37.3%
Thereof by Contiguity: 50.1% Thereof by Similarity: 4.1%
Recurrent associated meanings: chin, cheek, mouth, tooth, bone/skeleton, palate, gums

'Jaw' (sometimes 'lower jaw' specifically) is colexified with 'chin' in many languages in the sample (see also Buck 1949: 220 for relations between 'jaw,' 'chin,' and 'cheek' as well as, more rarely, 'beard' in Indo-European), namely in Yoruba, Baruya (by the analyzable term *maanagɨnya* /maanga-yagɨnya/ 'mouth-bone'), Kaluli, Kwoma (by the term *teekibi*, perhaps containing *teeki* 'to string, tense' and *bi* 'point'), Kyaka, Lavukaleve, Mali, One, Southeastern Tasmanian, Toaripi, Kosarek Yale (where the relevant term colexifies "wattle of an agama"), Abzakh Adyghe (by the analyzable term *żepqˀ* /że-pqˀ(ə)/ 'mouth-skeleton'), Greek, Japanese, Ket (by a semianalyzable term containing *ìt* 'tooth;' such a term is also found in Embera), Khalkha, Nivkh, Welsh, Kolyma Yukaghir, Blackfoot, Upper Chehalis, Cheyenne, Chickasaw, Itzaj, Kiliwa (by a term containing *yu=wha?* 'face,' which is literally 'eye=mouth'), Nez Perce, Nuuchahnulth (by the analyzable term *hiiṅiiƛ̓aksuɬ* /hiiṅiiƛ̓a-ʔaksuɬ/ 'lower/below-at.the.mouth/at.the.lips,' colexifying also 'lower lip'), Pipil (by the analyzable term *-te:n-tsi:ka-w* '-mouth-ant-POSS'), Arabela, Aymara, Cashinahua, Embera (where the meanings are associated with different genders), Guaraní, Jarawara, Macaguán (by a semianalyzable term containing *-bʉk* 'heel'), Rama (by a semianalyzable term containing an element meaning 'tree' and colexifying 'grater'), Tehuelche, Yanomámi, Mandarin, White Hmong, Samoan (by the analyzable term *'au-vae* 'CLASS.LONG.OR.NARROW.THINGS-foot/leg,' colexifying also 'foot of hill'), Takia (by the analyzable term *awa-n to-n* 'mouth-3SG arm-3SG'), Tetun, and Yay. Moreover, Katcha has *ku'ba mo buruŋɛ* 'bone GEN chin,' Waris *keu-mul* 'chin-edge,' Carib *-ekusalɨ yepo* '-chin POSS,' Cubeo *yedʉ-cũ* 'chin/jaw-CLASS.ROUND.ON.ONE.SIDE.AND.PLAIN.ON.THE.OTHER,' and Hawaiian *papa 'auwae* 'flat.surface chin' for 'lower jaw' specifically.

Alongside the complex or semianalyzable terms containing an element meaning 'mouth,' this association is also present in San Mateo del Mar Huave *mi-machat ombeayaran*, analyzable as 'AL.POSS-machete mouth,' and the situation in Baruya, with elements meaning 'mouth' and 'bone,' is precisely mirrored in San Lucas Quiaviní Zapotec and Bororo. Semianalyzable terms with an element meaning 'mouth' are alongside Kiliwa also found in Sora, Santiago Mexquititlan Otomí, and Xicotepec de Juárez Totonac. In Sora, moreover, *tˀo:dən ~ 'tˀo:dən ~ 'tˀu:dən* 'jaw' is colexified with 'mouth of animal or vessel' more specifically, and Muna *ghongki* can also refer to the "inner part of the mouth." A lexical tie with

'bone' is, alongside the languages mentioned so far, also found in Tsafiki (*te'fún chide* 'tooth bone;' this term colexifies 'gums,' as is the case in Upper Chehalis) and Wichí, which has *toch'ay lhile*, containing *ch'a* 'cheek' and *lhile* 'bone.' A term with such structure is also found in Kiliwa, and there is a semianalyzable term in Carrier. As for 'tooth' as a constituent, Hawaiian has *papa niho* 'flat.surface tooth,' colexifying "row of teeth, set of teeth" as well as "bridge for false teeth," and as mentioned above, there are semianalyzable terms in Ket and Embera. Moreover, Ngambay and Wappo colexify 'jaw' with 'molar tooth' (and diachronic development from 'jaw' to 'back tooth' is attested in Romanian, Buck 1949: 221), and further analyzable terms of the lexical type for 'jaw' involving a constituent meaning 'tooth' are found in Hupda (*təg-cáʔ* 'tooth-box,' for which compare Spanish *quijada*, derived historically from Latin *capsa, capsus* 'box,' Buck 1949: 221), and Fijian (*kau-ni-bati* 'stick/wood/tree-POSS-tooth'). Cubeo has a derived term.

The Kiliwa and Wichí association with 'cheek' is mirrored in Buli, Yoruba, Gurindji, Yir Yoront, Welsh, Lesser Antillean Creole French, Wintu, Abipón, Miskito, Kapingamarangi (by the analyzable term *gau wae* 'side foot/leg'), and Manange by colexification, and in Yir Yoront by the term *wal-kur* 'cheek/temple-slashing.sword.' Furthermore, Rendille *gi'daámme* is grammatically the plural of *gi'dáam* 'cheek,' and in Basque, the meaning of *baraila* varies dialectally between 'jaw' and 'cheek' (it can also assume the meaning 'revelry, fray'), and for Hani *baqxoq* 'jaw' (perhaps related to *xoq* 'lock'), compare *baqba* 'cheek.'

Finally, Dongolese Nubian and Nez Perce colexify 'jaw' with 'palate' (the relevant Dongolese Nubian term is also the name of a village).

Other associations include: Efik *mbañ* colexifies 'jaw' with 'gill of fish' as well as, figuratively, "[l]oud, impertinent talking." Hausa *ma-k'aru* is analyzable as 'LOC-finish.' This term is primarily applied with reference to fish, and also colexifies "top, inside, of a cooking-pot" and other vessels. Noni *kediɛw* consists of *diɛw* 'word, language, noise, sound' and the noun class marker *ke-*. One *sesu* also means 'to squish.' Yir Yoront *thamanwalq* contains *man* 'throat, neck' and *walq* 'hollow place,' while Central Yup'ik *agluquq* is analyzable as /agluq-quq/ 'center.beam.of.a.structure-one.that.is.' Miskito colexifies 'jaw' with 'beard,' Toba with 'operculum,' and Hawaiian with 'cheekbone' as well as "to talk a lot, jabber" and culture-specific instruments. Lenakel *kauga* also means 'corner,' Tetun *hasan* also means 'ambush,' as well as "notch in wooden columns of buildings for the positioning of beams."

119. The Kidneys

Representation: 84%
Motivated: 24.6%
Thereof Analyzable: 17.1% Thereof Colexifying: 7.5%
Thereof by Contiguity: 2.3% Thereof by Similarity: 13.5%
Recurrent associated meanings: back/behind, heart, seed, fruit, liver, bean, testicle, stone, fat, urine/urinate

The 'kidneys' are often named by terms making reference to their particular shape. Common are comparisons with fruits of similar shape, either by colexification or by morpho-

logical complexity. In Ngambay, *mùnjù* also means 'bean' (a pattern of colexification shared with Kiowa, where it can also refer to a 'pea,' and mirrored by the complex Chayahuita term *maquira-pi* 'bean-CLASS.FRUIT.BODY.OR.BACK'), in Mbum, *ánjòkò háì* consists of *ánjòkò* 'peanut' and *hàì* 'flesh,' and in One, *wasou tala* contains *wasou* 'betel nut' and *tala* which either means 'seed' or 'round thing' more generally. In Kapingamarangi, the word for 'kidney,' *ibu mimi*, contains *ibu* 'coconut shell' (which also means 'cup,' presumably by functional contiguity) and *mimi* 'urine.' In Yuki, *pohoṭ ~ pŏhōt* 'kidney' may be identical with *pohut ~ po´hoṭ* 'oak galls' and in Wappo, spoken in the vicinity and perhaps genealogically related, *ʔá·we* also may refer to potatoes or "any variety of food bulbs and corms." Fijian *ivi* is also the name of the native chestnut tree, and Nunggubuyu *wudu* is also the name of a tree with kidney-shaped nuts.

Frequent are also associations with 'fruit:' Kwoma has *magiir wuwu siik* 'back swell.up fruit/seed,' Toaripi *kōu uti fare* 'back bone fruit,' Hawaiian *kōnāhua*, perhaps containing *konā* 'hard' and *hua*, meaning 'fruit,' 'seed,' and 'egg' inter alia. (this term colexifies 'grease' and 'fat,' for this, compare Samoan *fatu-ga'o* 'heart/seed-fat/lard'), Tetun *fuan-sorin* 'heart/fruit-side,' and semianalyzable terms are found in Piro (where the relevant term also means 'seed' and acts as a diminutive marker; in Koyraboro Senni, there is a dialectal variant which is identical structurally with a verb meaning 'to sow, plant, raise'), Bwe Karen (where 'fruit' and 'round object are colexified') and Lenakel (where this term colexifies 'side of ankle' and 'earlobe'); in Yay, the word for kidney, *maak²yaaw¹*, includes the classifier for fruits, *maak²* and in Hani, *hhoqsiq* contains the classifier for round things *siq* (*hhoq* means 'to wear clothes, be covered' inter alia). As for 'seed,' Kwoma has *magiir wuwu siik* 'back swell.up fruit/seed,' One *wasou tala* 'betel.nut seed/round.thing,' Jarawara *siro noki/noko ~ kasiro noki/noko* 'frog/tree.species eye/face/seed/color' (note that there is a semianalyzable term with the identifiable constituent meaning 'eye' in Kildin Saami), Hawaiian *kōnāhua*, perhaps containing *konā* 'hard' and *hua*, meaning 'fruit,' 'seed,' and 'egg,' and Samoan *fatu-ga'o* 'heart/seed-fat/lard' (the constituents glossed as 'seed' have still other meanings in some languages).

Takia colexifies 'kidney' with 'stone,' Wichí has a derived term, Rama has *ngalíng kálup*, containing *ngalíng* 'stone' and *up* 'eye,' and Kyaka has a term term where the identifiable constituent means 'hard' and 'pebble' alongside other things.

As evidenced by the One and Bwe Karen cases, at times 'round object' in general is associated with 'kidney.' This is also the case in Lengua, where a semianalyzable term exists with the identifiable cosnstituent meaning 'to be round.' A term which also appears to make reference to the shape of the kidney by way of metaphor is found in Arabela, where *cajiniajajau* appears to contain *cajinia*, the word for a sitting person or a baby that is old enough to sit and the classifier *-jajau* for round objects. On a more abstract level, terms for the kidneys making reference to their shape are found in Abzakh Adyghe, which has *čʔentʼəʔʷ* /čʔ(e)-n-tʔ(e)-ʔʷ(e)/ 'end-EPEN-poke.out/bend-pointed' and Pawnee, where *spiruusuʔ* probably contains underlying *as-* 'foot,' *piruus* 'crooked' and the nominal suffix *-uʔ*.

There are also languages in which the kidneys are conceptualized via their position in the body. This is found in two languages of New Guinea, Kwoma and Toaripi, as

seen above. In four languages in the sample, Hausa, Koyraboro Senni, Muna, and Abzakh Adyghe, 'back' and 'kidney' are associated by colexification ('lower part of the back' in particular in Muna). In Tetun, which has several words for the kidney, *fuansorin*, consisting of *fuan* 'heart, fruit' and *sorin* 'side,' is encountered. The Tetun example provides the transition of the discussion to connections between the 'kidneys' and other internal organs of the body. Most frequent are indeed those with the 'heart,' occurring in six sample languages, namely Ngambay, Guaraní, Yanomámi (where the relevant term may also be used to refer to other spherical objects) by colexification, and in Tetun, Samoan, and Imbabura Quechua (*yana shungu* 'black heart') by analyzable terms. Three languages, Badaga, Kiowa, and Ancash Quechua, colexify 'kidney' and 'liver' (Badaga also 'lungs' and 'larynx,' and Kiowa has several complex terms for both meanings available for disambiguation), and in Yanomámi, *amoyõri* also denotes the 'gall bladder.' A body-part metaphor is found in Khalkha and Rotuman, where 'kidney' and 'testicle' are colexified, while in Ancash Quechua, the 'kidney' is also called *ruru-n* 'egg/testicle-3SG' (see Jóhannesson 1949: 107 for Indo-European parallels). In Khoekhoe, *!nāb* is a semantically very general term that can be used to refer to 'offals' generally, including virtually all internal organs of the trunk.

Less frequent are terms for the kidney in which their function plays a role. Already mentioned was Kapingamarangi *ibu mimi* containing *mimi* 'urine;' in Sedang, *bong núm* appears to be analyzable as /bông núm/ 'white urinate,' and there is a semianalyzable term featuring an element meaning 'urine' in Toba.

Other associations include: Buli *yiini* also means 'to live on others, be greedy,' the Efik term *ek'put* is also used inter alia to refer to "the tassels on the neck of a goat," and Hausa *k'oda* is also a verb meaning "[r]epair and sharpen the edge of a tool by beating." The Burarra term *-gurday* is lexically connected to *gu-day* 'friendship,' and Dadibi *ene nawe* appears to contains *ene* 'arrow, rattan.' Meyah *ofómfúf* appears to contain *ofóm*, which can mean 'ripe' or 'root.' Sahu *gogolatíla* might contain *gogo* 'body hair, fur, feather,' and Sko *háng* also means 'end of intestine.' Yir Yoront *man-nerp* is analyzable as 'neck-spirit.child,' and *mortworrqworr* contains *mort* 'pile' and *worrqo* 'big.' Kolyma Yukaghir *mumul* also means 'fist,' while the Cahuilla term *pípiviskun* is etymologizable as /pípivis-kunil/ 'vomit-sack.' Kiliwa *snpap* contains *pap* 'bundle.' The Yaqui term *sikupuriam* contains *siiku* 'navel,' while San Lucas Quiaviní Zapotec *x:quèe'ts* also means 'gizzard.' Cavineña *epecaca* appears to be analyzable as /e-pere-caca/ 'INAL.POSS-rib-small,' and Cubeo *joedu* shares the root *joe* with a term denoting a tree species and an 'axe,' with different classifiers differentiating between the meanings. Hawaiian *pu'u-pa'a* is analyzable as 'mound-firm.' The term also means 'virgin, virginity' and is used figuratively to refer to emotions. Mandarin *yao1* also means 'hip,' Rotuman *ififi* also "in bunches or clusters," and Yay *yaaw*^{1} also 'monster, ogre.'

120. The Lip

Representation: 94%
Motivated: 47.4%
Thereof Analyzable: 23.5% Thereof Colexifying: 24.6%
Thereof by Contiguity: 16.1% Thereof by Similarity: 8.9%
Recurrent associated meanings: mouth, edge, skin, beak, peel/rind/shell, bark, snout, female genitalia, entrance/door, language, cover, leather/hide, leaf

When words for the 'lips' are analyzable, the most frequent pattern is that the respective words consist of those for 'mouth' and 'skin' (with 'skin' at times having related meanings), as in San Lucas Quiaviní Zapotec *gui'dy ru'uh* /gui'ihdy ru'uh/ 'skin mouth.' This is found in nineteen of the sampled languages, next to San Lucas Quiaviní Zapotec in Buli (where an additional element meaning 'leaf' is present; the whole term also can refer to the "notch of a flute"), Efik, Mbum, Anggor, Kwoma, One, Toaripi, Yei, Sora, Highland Chontal, Kashaya, Lakhota, Xicotepec de Juárez Totonac, Cavineña, Cayapa, Kaingang, Maxakalí (where 'skin' colexifies 'cover,' compare also Yana *sob?liyauwal(la)* /sab?li-yauwal(la)/ 'cover-NMLZ-mouth'), and Yay, and thus occurs everywhere except Eurasia (the pattern is also etymologically detectable for Nunggubuyu); there is also an optional compound in Bislama of this type to disambiguate a term meaning both 'lip' and 'mouth' and semianalyzable terms are found in Sko, Haida, and Xicotepec de Juárez Totonac. Rama has *sík úk* 'tooth skin' (the language features also another semianalyzable term where the identifiable constituent is 'skin'). Since 'skin' is lexically associated with several related meanings (see discussion in section 135), there are also secondary associations in many of the abovementioned terms. This is the case with 'bark' in Efik, Kwoma, Toaripi, Xicotepec de Juárez Totonac, Kaingang, and Maxakalí, with 'rind,' 'peel,' or 'shell' in Efik, Lakhota, Xicotepec de Juárez Totonac, Cayapa, Kaingang, Maxakalí, and Rama, with 'leather' in Sora and Cavineña, and with 'cover' in Lakhota. There is a seminalyzable term with the identifiable constituent colexifying 'skin' with 'bark' and 'covering' in Sko.

There is also a wealth of complex terms in other languages where an element meaning 'mouth' is present, but the other consituent does not mean 'skin': one relatively frequent variant of the complex terms of the 'mouth-skin' type is to have 'mouth-edge,' which is for instance found in Kiliwa (*ha?hiiy* 'lips,' *ha?* 'mouth' *hiiy* 'edge, border;' this term is itself glossed as 'border;' the language also has the term *ha?=nat-u?=kw+waa* 'mouth=top/atop-OBL-WH+sit' for the 'upper lip' specifically) as well as in Toaripi, Sora, and Fijian. When an association with 'edge' occurs, however, this is more frequent by colexification, which is found in sixteen languages, namely Ngambay, Rendille, Muna (also colexifying 'side'), Sahu, Basque (also colexifying 'corner'), Welsh, Itzaj, Bora, Bororo, Carib (where 'upper lip' is the colexified meaning more specifically), Cashinahua, Chayahuita, Embera, Guaraní, Miskito, Piro, Yanomámi, and Malagasy; furthermore, Yanomámi has *kasiki*, consisting of *kasi* 'edge' and the quantal classifier *ki* (see § 4.4.1.1.). In two languages, Buli and Samoan, terms for 'lip' are found which consist of the respective words for 'mouth' and 'leaf;' Hausa also has an association by colexification with 'leaf bud.' Further complex terms involving a constituent meaning 'mouth' are Efik *mfut inua* 'shade mouth,'

Kanuri *kâ-cî-bè* 'stick-mouth-of,' Khoekhoe *nino a'dɔgɔ* 'mouth above' for the 'upper lip' specifically, San Mateo del Mar Huave *apal ombeayaran* 'lid mouth,' Pawnee *haaka-huukita-haahiriˀ*, analyzable as 'mouth-on.top-LOC,' also for the 'upper lip' specifically, Yaqui *tenberia*, analyzable as /teeni-be'eri-a/ 'mouth-exceed-NMLZ,' Copainalá Zoque *aŋsis* /aŋnaca-sis/ 'mouth-flesh,' Arabela *rupaa-que* 'mouth-CLASS.CLOTH,' Cubeo *jije-tarabʉ* 'lip/mouth-CLASS.CIRCULAR,' Kapingamarangi *malau ngudu* 'two.things.joined.together mouth,' and Lenakel has *nɨvhivhi-nhul* 'end/tip-mouth' for the 'upper lip' and *netpɨ-nhul-* 'belly-mouth-' for the 'lower lip.' In Ngambay, Swahili, Burarra (where there is a very general term referring to the entire mouth area, including 'cheek' and 'chin'), Kyaka, Acoma, Cheyenne, Comanche, Itzaj, Nez Perce, Nuuchahnulth, Oneida, Wintu, Bororo, Huambisa, Imbabura Quechua, Bislama, and Tetun, 'lip' and 'mouth' are directly colexified (Sahu colexifies 'lip' with "the outside of the mouth," and Nez Perce also with 'mouth of river' and 'cave;' this association is frequent in Indo-European, Buck 1949: 229-230). Semianalyzable terms where one constituent is 'mouth' are found in Katcha, Khoekhoe, Abzakh Adyghe, Japanese, Kolyma Yukaghir, Biloxi, Chickasaw, Wappo, Aguaruna, Hupda, and Rotuman. Mostly due to colexification with 'mouth,' the 'lips' are also associated with 'language,' which is the case in Wintu, Imbabura Quechua, and Hawaiian; the only language including a term which colexifies 'lip' and 'language,' but not 'mouth' is Hawaiian.

In Swahili, Kyaka, Carrier, Central Yup'ik, Jarawara, Yanomámi, and Takia, 'lip' is colexified with 'beak' (see also section 5; in Kyaka with "wide beak," in Central Yup'ik with "upper or lower part of a snout or beak" and in Takia with 'pounting lips' more specifically). The Pipil and Jarawara terms colexify 'lip' with 'snout' (in Jarawara alongside 'fruit' and 'mass, hunk, lump, whole'), while in Acoma, 'upper lip' more specifically is colexified with 'snout' and conversely, in Central Yupi'k 'upper part of snout' is colexified with 'beak.' In four languages, Ngambay, Burarra, Kyaka, and Macaguán, the word for 'lip' is extended to also mean 'entrance' or 'door;' however only in Macaguán does the respective term not also denote the 'mouth' at the same time, a meaning for which this metaphorical extension is more common (see section 124 and § 6.2.3.2.). In four of the sampled languages, Ngaanyatjarra, Lake Miwok, Bislama, and Hawaiian, there are extensions to the female genitalia (as in Latin; in Hawaiian among other meanings colexified). Rarer metaphorical extensions are found in Nivkh, where *ŋapu* also means 'bay,' in Central Yup'ik, where *qerrluq* may refer to 'stones around a firepit,' and in Cashinahua, where *kebichi* (given as *kebixi* in the Spanish-Cashinahua section of the consulted source) may also be used with the meaning 'eyelid.' Finally, in five of the sampled languages, 'upper lip' and 'lower lip' are expressed by different lexical items. These languages are Toba, Lenakel, Dadibi, Lavukaleve, and Katcha.

Still other associations are: Efik *dum'baru* colexifies "[f]irst young leaf buds of any tree," "[g]erminating of seeds," and "[f]irst coming through of a tooth." Kyaka *kambu* also means 'vote, voice' (likely due to Kyaka colexifying 'lip' with 'mouth'), Meyah *ofúj ofóu* contains *ofóu* which can mean either 'egg,' 'meaning,' or 'dust,' and Waris *muemb-ta* appears to be analyzable as 'saliva-small.object.' Itzaj *chi'* is also the name of a tree, while Nez Perce *túːskiɬikuː siṕéːwn* 'upper lip' is analyzable as 'uppermost twist-ACT.PTCPL.' Central Yup'ik *qerrluq* also means 'stones around a firepit,' and Wayampi *ɛmɛ* also 'kerf in flute,'

'fabric edge of hammock rope,' and 'shore.' Hani *meiqjil* contains *jil*, which inter alia means 'to finish something' (and for *meiq*, compare *meiqdaoq* 'mouth, snout'?). Hawaiian *lehe* can also mean 'stretched' or 'loose' and also denotes a deep-sea fish.

121. The Liver

Representation: 93%
Motivated: 13.4%
Thereof Analyzable: 1.5% Thereof Colexifying: 11.9%
Thereof by Contiguity: 0.0% Thereof by Similarity: 4.7%
Recurrent associated meanings: soul, heart, lungs, core/center/middle, belly/stomach, kidney, chest

There are relatively few lexico-semantic associations concerning the 'liver;' the ones that are attested link it mostly to other internal organs of the body. In five sampled languages, Ngambay, Kwoma, Sahu, San Lucas Quiaviní Zapotec, and Yay, colexification of 'liver' and 'heart' is found, in Imbabura Quechua, the 'liver' and the 'kidneys' are called *yana shunga* 'black heart' (in Irish, 'liver' is a compound of terms meaning 'heavy' and 'heart,' Buck 1949: 252). In Sahu, the simplex *katere* has rather broad reference, including, alongside 'liver,' 'chest,' 'heart,' and 'soul' ('chest' and 'liver' are also colexified in Efik). To single out the meaning 'liver,' *katere ma bibiwisi* may be used; *bibiwisi* is a reduplicated version of *biwisi* 'banana flower,' which has a reddish color similar to that of the liver. Other languages (Badaga, Kiowa, and Imbabura Quechua) colexify 'liver' and 'kidney,' and in still others, (Kwoma, Badaga, Laz, Cahuilla, and Yay), 'liver' and 'lungs' may be referred to by the same lexical item (this association may also be present in Anggor, though marked as dubious in the source). In the case of three languages, Ngambay, Khalkha, and Yanomámi, terms for 'liver' also include meanings such as 'belly,' 'tummy,' or 'stomach.' In Khoekhoe, there is a term covering internal organs ('lungs,' 'heart,' 'liver,' and 'kidneys') which is also capable of referring to the 'stomach.'

Relatedly, in Central Yup'ik, *tenguk* also means 'solar plexus,' and in Badaga, *i:raḷu* also denotes the 'entrails' and 'intestines.' In four languages of the Americas, Highland Chontal, Wappo, San Lucas Quiaviní Zapotec, and Yanomámi, the words for 'liver' also have a more abstract reading of 'core,' 'center,' or 'middle,' and Efik *ë'sët* inter alia also denotes the 'inside, interior' of something (Bowden 1992: 36, as cited by Heine and Kuteva 2002: 199, reports a grammaticalization path 'liver' > 'in' in a number of Oceanic languages).

The liver is also the seat of emotions or the soul in a number of languages (or more appropriately put, conceived of this way in the associated cultures). This is noted for the African languages Efik and Noni, and is also found in New Guinea in Kwoma, Meyah (see examples in § 4.5.1.4.1.), Sahu, Sko, and Toaripi, as well as in White Hmong in Southeast Asia. In addition, in Fijian, 'courage' and 'cowardice' are associated with the 'liver,' Hawaiian *ake* is ambiguous between the nominal reading 'liver' and a verbal reading that has to do with desiring or yearning for something; optionally, *ake-pa'a* 'liver-firm' may be used to single out the nominal reading specifically, while in Embera, the sequence *dadyí* in

dadyíthãrí 'liver' occurs in many expressions having to do with human beings, their souls and their spirit.

Other associations include: Hausa *hanta* is also used figuratively for "one's dearest possession," Koyraboro Senni *tasa* also denotes a 'metal bowl' and means 'to push' as a verb, while Ngambay *wùr* also means 'patience.' Baruya colexifies 'sorrow,' and Kyaka *pungi* also means 'rope.' One *wala* is also used with the meaning 'side,' Meyah *odóu* also means 'front,' and Sahu *katere* also 'chest.' Sko *pung* is also the name of a large bamboo species and also means 'to butcher, cut up meat.' Kosarek Yale *bubu* also means "carrying or bearing part, point of support." Abzakh Adyghe *sʷʔə* also means 'good, useful' (though it occurs with the meaning 'liver' only in a redundant complex term). Badaga *cuṭṭage ~suṭṭage* also denotes the 'larynx,' and Basque *gibel* also means 'listnessness, lethargy.' Khalkha *elige* figuratively also denotes the 'breast' and 'blood relatives.' The Sora term *ɔ'gare:n* is also used to denote "pulp of fruits, pith, kernels etc.," and Welsh *iau* also means 'yoke' and 'Thursday.' Highland Chontal *łada* is also glossed as 'center of being.' Pipil (Cuisnahuat dialect) *-el-tapach* is analyzable as '-inside-shell,' and Maxakalí *xupkũnãy* might contain *xup* 'to suck, sip, hang down.' Hani *caoq* also inter alia means 'to contain, to be present inside,' 'a frightening place where spirits live,' and acts as a classifier for clumps or clusters of things. Mandarin *gan1* also means 'dry' (in both readings reflecting Early Middle Chinese *kan*, Pulleyblank 1991: 102), and Rotuman *äfe* also 'thousand' inter alia. Takia *ate-* is also used with the meanings 'palm of hand' and 'plain among hills,' and White Hmong *siab* also means 'high, tall.' Lesser Antillean Creole French *fwa* also means 'time, turn, occasion' and 'faith.'

122. The Lungs

Representation: 93%
Motivated: 25.4%
Thereof Analyzable: 16.5% Thereof Colexifying: 8.9%
Thereof by Contiguity: 7.9% Thereof by Similarity: 5.8%
Recurrent associated meanings: liver, light, heart, foam, chest, float, air

Words for the 'lungs' are often associated with other internal organs of the body lexically, most often with the 'liver.' Kwoma, Badaga, Laz, Cahuilla, and Yay colexify 'lungs' and 'liver' (Badaga also 'kidney' and 'larynx;' perhaps Anggor does, too, but this is marked with a question mark in the source), and there are complex terms in Ket (*mensiŋ* /ben-sēŋ/ 'apart-liver/innards'), Hupda (*hɔ wowŏw'* 'liver wring.out'), Ancash Quechua (*yuraq ñatin* 'white liver'), Bislama (*waet-leva* 'white/bright-liver'), Fijian (*yate-vuso* 'liver-foam'), Hawaiian (*ake-māmā* 'liver-light,' *ake-makani* 'liver-wind,' and *ake-pāhola* 'liver spread'), Kapingamarangi (*ade di baahi* 'liver ART side'), and Tetun (*aten-book* 'liver-move'). Note also Yoruba *ẹ̀dọ̀-fóro* 'liver-lung.' There is a semianalyzable term in San Lucas Quiaviní Zapotec, where the identifiable constituent colexifies 'liver' and 'heart,' and it is with the latter organ that 'lungs' is also often associatied in the sampled languages. Buin, Burarra (by the analyzable term *mun-molma* 'CLASS.DOMESTIC-warmth'), Kwoma, and Yay have terms capable of referring to both organs (and they are considered the seat of emotions in Buin and Bu-

rarra). Guaraní has *ñe'â vevúi* 'heart light' (though *vevúi* alone can refer to the 'lungs,' too), and Ancash and Imbabura Quechua a term literally translatable as 'white heart' (*yuraq shunqu* and *yurak shungu* respectively). Furthermore, there is a semianalyzable term with the identifiable constituent meaning 'heart, breast' in Kaingang. Moreover, in Khoekhoe, there is a general term for 'offals' which can refer to virtually all internal organs of the trunk, while there are other lexical items for each one specifically.

Almost all other recurrent lexico-semantic associations already occur in the terms mentioned so far, making reference both to the light weight and the spongy structure of the lungs that sets them apart from other internal organs of the body. Alongside Guaraní and Hawaiian, terms where their light weight is used for denomination are Yir Yoront *ngerr-lolt* 'belly-light.in.weight/hollow' (the source notes that this term is "[g]iven fairly consistently also for 'heart'"), Welsh *ysgyfaint* /ysgafn-aint/ 'light-ABSTR,' Chickasaw *i̲shoppaya,'* analyzable as /im-shoppaya-'/ 'DAT-be.light-NMLZ,' Arabela *pa namaca* 'our light,' and 'lungs' and '(be) light' or 'something light' are colexified also in Nuuchahnulth, Bora, and Samoan (note in addition the similarity between Tsafiki *que'fó* 'lungs' and *que'fún* 'light'). This association is also recoverable etymologically in German (and Germanic more broadly, Kluge 2002). Kluge has it that "die Lungen werden als 'die Leichten' bezeichnet, weil Lungen von Schlachttieren als einzige Innereien auf dem Wasser schwimmen" / "the lungs are called 'the light ones' because lungs of animals for slaughter as the only innards float on water," see also Jóhanneson (1949: 105) for Indo-European more broadly. The 'lungs' are colexifed with 'foam' in Buin, Kaluli, and Toaripi, and a complex term is, alongside Fijian, found in Chayahuita *sa'poro'* (/sa'po'-ro') 'foam-CLASS.PILE.' Similarly, Guaraní colexifies 'lungs' with 'be foamy, sudsy,' Jarawara has *hasa-bori* 'make.bubbles-container,' and Wintu *λoso* denotes both 'lungs' and 'foamy crest on waves.'

Consistent with Kluge's explanation, Nuuchahnulth also colexifies 'to float' and Central Yup'ik has *pugtaun* /pugta-(u)n/ 'float-device.for;' this term colexifies 'lungs' with 'float' (in the sense of 'raft') as well as 'buoy' and 'life-vest.' Similar to the association with 'wind' in Hawaiian mentioned earlier, the fact that the lungs are used for breathing is mirrored by associations with 'air' in Oneida (*yewelalákhwaʔ*, analyzable as /ye-wel-l-hkw-aʔ/ 'FEM.INDEF.SG.AGENT-air/wind-be.in-INSTR-HAB') and Wichí, which has a very similar term; likewise, the meanings 'lung' and 'breath' are connected diachronically in Greek. Tasmanian (Middle-Eastern, Southeastern, and Western) and Lesser Antillean Creole French colexify 'lungs' with 'chest.' There are also associations with 'meat' and/or 'flesh,' however, only by semianalyzable terms, namely in Mbum, Abzakh Adyghe, and the Norton Sound - Upaliq dialect of Central Yup'ik.

Other associations include: Hausa *huhu* is also the name of a lung-related disease of horses (and has still further meanings). Koyraboro Senni *kumbu* also dialectally denotes a 'type of hoe,' and Baruya *pawajɨkawaai ~ pawajɨkawaalo* contains *pawajɨka* 'spleen.' Buin *turupa* also means 'saliva' (compare the association between 'foam' and 'saliva' described in section 24; Buin is among the languages colexifying 'lungs' with 'foam'). Dadibi *ogwa bai*, also meaning 'spirit,' is analyzable as 'son baby,' Muna *ghau* also denotes "overgrown tubers/maize." Sahu *'aba'abala* appears to be reduplicated from *'abala*, meaning "(wall made of) midribs of sago palm fonds." Badaga *cuṭṭage ~ suṭṭage* also is used for 'kidney,' and Haida

hlgucu also denotes a 'fish's swim bladder.' Itzaj *b'oj* also means 'to knock' inter alia. The Kiliwa term *x-?+kwiiy-waa-u?* is semianalyzable as 'CAUS-DN+??-sit-OBL' (compare *-kwiy* 'cloud'?), and Kiowa *k'αhyoudl* is related to *k'αe* 'skin' and *-houdl* 'intensive.' Wintu *ẋ0s* 'fog, steam, gas, lungs' yields *ẋoso* 'lungs.' Copainalá Zoque *se'u* also means 'to have a sore throat.' Cayapa *jenana* might contain *nana* 'balsa' (notably, a particularly light kind of wood), and Cubeo *vɯibo* consists of *vɯi* 'tuber' and the classifier for round or hard objects *-bo*. Guaraní *tajygue* also means 'tendon, vein, nerve, muscle' and 'strength,' Kaingang *fe kãnhvy* also means 'joy,' while the Rama term *ikúngkungma* appears to be based on *kung* 'louse, whit, air root.' Toba colexifies 'lung' with 'tears,' Wayampi *tuluɛwɨy* contains *tulu* 'big,' and Yanomámi *herekɨ ~ hẽrẽkɨ ~ heremopɨ* contains *here* 'wet' and a quantal classifier (see § 4.4.1.1.). Finally, Hani *povq* also means 'to get soft' and 'to be empty, used up,' Bwe Karen *θə́'ó* also means 'to weed,' and Rotuman *ma'ma'a* is reduplicated from *ma'a*, which means, inter alia, 'light in color.'

123. The Milk

Representation: 91%
Motivated: 54.7%
Thereof Analyzable: 23.9% Thereof Colexifying: 30.8%
Thereof by Contiguity: 48.9% Thereof by Similarity: 2.4%
Recurrent associated meanings: breast, water/liquid/juice, udder/teat, nipple/teat, suck, resin, semen, pap, Burton's legless lizard, cow

There is one cross-linguistically dominant association in terms for 'milk' that is unsurprising, namely that with 'breast.' This is frequently realized by analyzable terms (40 languages), in which case 'milk' is canonically literally 'breast water' (e.g. Sentani *nimə bu*) or, more generally 'breast liquid' (e.g. Bororo *mokuro kuru*) or 'breast juice' (e.g. Kaluli *bo ib* /bó ib/). Such terms are also found in Efik, Muna, Waris, Ket, Kiliwa, Kiowa, Pawnee, Cavineña, Embera, Guaraní, Hupda, Jarawara, Lengua, Miskito, Piro, Tsafiki, Wayampi, Yanomámi, Great Andamanese, Hani (where *aqqul* colexifies 'sweet' and indeed likely also contains *qul* 'sweet'), Bwe Karen, and Malagasy; Buli and Kapingamarangi have semianalyzable terms where the identifiable constituent means 'breast.' In languages with systems of nominal classification, as already seen in § 4.4.1, the respective terms typically involve a classifier affix, as is the case in Arabela (where 'breast' and 'teat' are colexified), Bora, and Chayahuita; similarly, in Swahili, the same root yields both meanings depending on noun class assignment. Variants of the morphologically complex terms mentioned above are Meyah *méngk ofód* 'breast flood' and San Mateo del Mar Huave *aonts mijiwaran* 'excrete breast,' and there are semianalyzable terms with 'water' or 'juice' in Cayapa, Rama, and Great Andamanese.

A further variant are terms in which the consituent meaning 'breast' is replaced by a verbal element meaning 'to suck,' such as in Chickasaw *pishokchi'*, which is analyzable as /pishi okchi'/ 'suck juice/liquid.' The same pattern is also found in Rotokas (*roroo ovi* 'develop.breasts/breastfeed/suckle CLASS.LIQUID') and Tetun (*susu-been* 'suck-liquid'); however, Tetun also has a term of the derived type, namely *susu-n* 'suck-SINGULATIVE,' and in-

deed, derived terms with a similar structure are found in more sampled languages, namely in Khoekhoe, Kolyma Yukaghir, and Haida, and in Bislama and Samoan, *susu* is ambiguous as to part of speech and means 'milk' in nominal and 'to suck' in verbal usage. Similarly, Pawnee colexifies 'breast' with 'to suck,' and hence the relevant complex term for 'milk' also betrays an association with the latter meaning. Fijian *sucu* colexifies 'to be born, to suck the breast' with 'birth, milk' (there is a optional complex term with *wai* 'liquid' on the basis of *sucu*); the association with sucking is also attested weakly in Indo-European, the evidence being an Albanian word for 'cheese' containing a root attested in Sanskrit with the meaning 'suck' (Buck 1949: 385).

Colexification of 'milk' and 'breast' is even more frequent, meaning effectively that forty-nine languages have a single term to cover both 'milk' and '(female) breast:' these are Ngambay, Noni, Yoruba, Anggor, Baruya, Buin, Burarra, Gurindji, Kyaka, Lavukaleve, Ngaanyatjarra, Nunggubuyu, Toaripi (colexifying also 'scrotum' and denoting a particular breast-shaped shellfish), Sahu, Kosarek Yale, Basque, Kolyma Yukaghir, Carrier, Upper Chehalis, Cheyenne, Haida, Kiowa (colexifying also 'to flow, melt'), Lake Miwok, Nez Perce, Nuuchahnulth, Oneida, Santiago Mexquititlan Otomí, Quileute, Xicotepec de Juárez Totonac, Wappo, Wintu, Yuki, Central Yup'ik (Yukon and Norton Sound dialects), San Lucas Quiaviní Zapotec, Aguaruna, Aymara, Cashinahua, Huambisa, Kaingang, Macaguán, Bislama, Hani, Hawaiian (here there is the term *wai-ū* 'water-breast' which however can refer to 'breast' itself), Lenakel, White Hmong, Rotuman, Samoan, Takia, and Tetun. Biloxi may be another case of colexification. Here the source gives *wa´k tasi´* /wak tasi´/ 'cow female.breasts' as the term for milk (for which compare Maxakalí *mũnũytut yõktat hep* contains *mũnũytut* 'cow' and *hep* 'liquid'), which suggests that this in fact refers to milk as a foodstuff and in fact *tasi´* alone can refer to both 'milk' and 'breast.' Hausa, by another term, colexifies 'breast' and 'sour milk' inter alia. For Kashaya, it is noted that *šiʔdo* 'breast' is also the old term for milk (for which there is now *molokko*, a loan from Russian). Interesting in this context is the case of Wintu, which features a single term for both referents but also has another term, *wuh*, which denotes both 'cattle' and 'milk,' and this kind of provenience contiguity may have come into being analogously.

Associations with 'breast,' either by analyzability or colexification, are common all over the world, with the notable exception of Eurasia, in which the association is only found in Basque and Ket; it is also not reported by Buck (1949: 385) for Indo-European. Redundant (or seemingly redundant) terms are also common, such as Yir Yoront *thaynpolqor* 'breast-milk,' although such formations may either be motivated by the introduction of cow's milk as a foodstuff and serve to distinguish mother's milk from it or to single out readings of the simplex for 'milk' if this is itself colexifying, for instance with other liquids. For instance, Kwoma *pi* ranges semantically over 'blood,' 'sap,' and 'milk,' and this may motivate the presence of *muku pi* (*muku* is 'breast').

Further patterns of colexification, such as that with 'blood' and 'resin' in Kwoma (the association with 'resin' is also found in Sora and Ineseño Chumash) already mentioned above include that with 'milky-looking pus' in Kyaka, that with 'soup' in Rotokas, that with 'milky juice' and 'latex' in Basque, that with 'juice' in Sora and Ineseño Chumash,

that with 'semen' in San Lucas Quiaviní Zapotec and Bislama, and that with 'pap' in Yoruba and Tuscarora.

A further pattern of colexification, encountered in Yoruba, Basque, Kolyma Yukaghir, Lavukaleve, Ineseño Chumash, Haida, Lake Miwok, Nuuchahnulth, Wintu, Aguaruna, Aymara, is that with 'nipple, teat,' and it is likely due to the colexification with 'breast' and 'nipple' by spatial contiguity rather than a genuine association, although Nez Perce and Ineseño Chumash colexify 'milk' with 'nipple' while having a separate term 'for breast.' In Cubeo, the 'milk'-word *opeco* appears to be derived by unknown means from *ope* 'nipple,' though this remains somewhat unclear. The association with 'nipple, teat' is realized by analyzable terms in Arabela, Cavineña, Embera, Miskito, Wayampi, and Malagasy.

'Milk' is also colexified with 'udder, teat' in Hausa, Yoruba, Kyaka, Toaripi, Haida, Lake Miwok, Nez Perce, Nuuchahnulth, Tuscarora, Wintu, Aguaruna, Aymara, White Hmong, Rotuman, and Takia (which also colexifies 'to squeeze'), and associated with these meanings due to 'udder, teat' being a secondary meaning of 'breast' by analyzable terms in Efik, Arabela, Cavineña, Embera, Miskito, Wayampi, Yanomámi, and Malagasy.

Both Burarra and Nunggubuyu extend their respective terms for 'breast, milk' to also denote "Burton's legless lizard," a species of lizard native to Australia and New Guinea.

Still other associations include: Hausa *madara* is also the name of a "kind of native-made cloth," "English salt in loose form," "[t]he unboiled juice of certain fruits or of sugar-cane," "[u]nadulturated musc scent," as well as "[p]ure silver." Koyraboro Senni *waa* also means 'to defecate, shit,' and Yoruba *ọmú* is also the name of an "instrument used in weaving to divide the woof." Muna *susu* 'canned milk' (borrowed in this sense from Bahasa Indonesia) is also the name of a large shellfish inter alia. Cheyenne *nénééhe* also means 'bottle' (this term is register-specific). Lesser Antillean Creole French *let* also means 'letter, character,' Nuuchahnulth *ʔinma* also denotes "Nob Point, where white powder seems to run out of a breast-shaped rock," Wintu *ʔEm* also means "hold pectorally, carry something in the arms, embrace," Bislama *titi* and *susu* also mean 'unweaned,' Hawaiian *waiū* also is the term for a 'wet nurse' and *kea* also means 'white, clear' inter alia. Rotuman *susu* also means 'to sew.'

124. The Mouth

Representation: 97%
Motivated: 48.8%
Thereof Analyzable: 5.9% Thereof Colexifying: 42.9%
Thereof by Contiguity: 15.3% Thereof by Similarity: 6.6%
Recurrent associated meanings: opening, word/language/speech, beak, lip, edge/tip, door/entrance, muzzle/snout, estuary, tooth, hole, blade, riverbank, barrel of gun/muzzle of gun

Ngambay, Swahili, Burarra, Kyaka, Acoma, Cheyenne, Comanche, Itzaj, Nez Perce, Nuuchahnulth, Oneida, Wintu, Bororo, Huambisa, Imbabura Quechua, Bislama, and Tetun colexify 'mouth' with 'lip' (a common association by semantic shift in Indo-European,

Buck 1949: 228); furthermore, Aguirre Licht (1999: 99) mentions *í-tãe* 'lip-intersection' for Embera, but it is not entirely clear whether this term has lexical status, and Cubeo has a single root occurring in terms for both 'lip' and 'mouth,' with classifiers differentiating between the two.

Baruya, Gurindji, and Cahuilla colexify 'mouth' with 'tooth' or 'teeth,' and Jarawara has *inohoti/inohoti*, consisting of *ini/ino* 'tooth' and *hoti/hoto-ne* 'hole' (compare also Kosarek Yale *sikaan* 'mouth' and *si* 'tooth, thorn, point, name'). In fact, similar terms are found in other languages of South America: Maxakalí has *yĩy-kox* 'speak-hole' and Tsafiki *fi'quí foró* 'language opening/hole' (though note also that Lesser Antillean Creole French colexifies 'mouth' with 'hole' among other meanings directly). Furthermore, an element meaning 'tooth' may be diachronically detectable in the word for 'mouth' in Nunggubuyu. Similarly, 'mouth' is colexified with 'language,' 'word' and/or 'speech' in Efik, Hausa, Ngambay, Rendille, Muna, Ngaanyatjarra, Khalkha, San Mateo del Mar Huave, Kiliwa, Wintu, Arabela, Bora, Chayahuita, Guaraní, Miskito, Ancash and Imbabura Quechua, Yanomámi, and Sedang, which also colexifies 'knife' (and Baruya colexifies 'angry speech, anger' more specifically), and in Tasmanian (Middle-Eastern, Southeastern, and Western) terms for 'mouth' are clearly related to those for 'language, utterance, speak,' though the precise relationship is not recoverable. Consistent with the South American pattern in complex terms just mentioned, the Cubeo term for 'mouth' is differentiated from 'lip,' with which it shares its root, by a classifier for entities having to do with language.

There are, however, also many metaphor-driven extensions of 'mouth' to other meanings. Buli, Khoekhoe, Laz, Sora, Lesser Antillean Creole French, Itzaj, Yanomámi, and Mandarin colexify 'mouth' with 'muzzle' or 'snout' (a common pattern in Indo-European, Buck 1949: 228), Buli, Efik, Khoekhoe, Ngambay, Swahili, Baruya, Kyaka (colexifying 'wide beak' specifically), Toaripi, Abzakh Adyghe, Nivkh, Kashaya (where *haʔbo* is analyzable as /ʔaha-ʔbo/ 'mouth-enlarge/swell.up' and means both 'protrusion of the mouth' and 'external mouth'), Yaqui, Hupda, Fijian, Malagasy, Rotuman, Samoan, and Tetun with 'beak' (in Kyaka 'wide beak' specifically; similarly, Yir Yoront colexifies "bottom of bird's beak" inter alia), and Buli, Hausa, Kanuri, Khoekhoe, Koyraboro Senni, Baruya, Basque, Upper Chehalis, Itzaj, Santiago Mexquititlan Otomí, Pipil, San Lucas Quiaviní Zapotec, Arabela, and Bororo colexify 'mouth' with 'edge' and/or 'tip.' Similarly, Welsh colexifies 'mouth' with 'end' (and also with 'top' and 'head').

Efik, Khoekhoe, Toaripi, Abzakh Adyghe, Upper Chehalis, Nez Perce, and Mandarin colexify 'mouth' with 'estuary' (compare the derivational relationship between Latin *ōs* 'mouth' and *ōstium* 'door, entrance, river-mouth,' as well as a cognates of *ōs* in Old Prussian and Old Norse meaning 'river-mouth' according to Buck 1949: 228, and section 20 for evidence for this connection from the present sample). Nez Perce also colexifies 'mouth of cave.' Similarly, Efik *inua* also denotes an 'inlet, gap,' Lavukaleve *leu* also a "passage, channel in reef for a canoe to go," and Buli and Tetun colexify 'mouth' with 'riverbank.'

Buli, Efik, Hausa, Noni, Burarra, Kyaka, Ngaanyatjarra, Rotokas, Basque, Lengua, and Mandarin colexify 'mouth' with 'door' and/or 'entrance' (compare the Latin evidence quoted from Buck above). Dongolese Nubian, Rendille, and Basque colexify 'mouth' with

'blade,' and more specifically, Buin colexifies 'cutting edge of blade,' while Lake Miwok and Hawaiian colexify 'barrel of gun' or 'muzzle of gun.'

Furthermore, 'mouth' is extended to 'opening' in general (but often of bottles or other vessels in particular) in Efik, Hausa, Khoekhoe, Ngambay, Dongolese Nubian, Yoruba, Kyaka, Muna, Ngaanyatjarra, Rotokas, Toaripi, Abzakh Adyghe, Basque, Khalkha, Sora, Welsh, Lesser Antillean Creole French, San Lucas Quiaviní Zapotec, Bororo (also with 'curve'), Lengua, Miskito, Toba, Yanomámi, Bislama, Hawaiian, Lenakel, Mandarin, Rotuman, and Takia.

Other associations include: Buli *noai* has additional readings as 'advice,' 'command,' 'oath,' and others, while *takabi* means 'sherd, piece of clay vessel' and is a vulgar and insulting term for 'mouth.' Efik *inua*, similarly, colexifies 'mouth' with 'report' and 'boasting' inter alia. Hausa *baka* also means 'bow' and 'catch of lock,' again inter alia, and *baki* is glossed also as "conclusion, maturity." Koyraboro Senni *mee ~ miñe* also means 'loose' inter alia and Ngambay colexifies 'suck' and 'embrace,' again among other meanings. Kyaka *kambu* also means 'vote, voice.' The Rotokas term *gisipo* also means 'talk, words,' while *akuta* also means 'to open one's mouth, to shout' as a verb. Abzakh Adyghe *ʔᵊə* also means 'to speak,' 'beg' as well as 'to listen, understand' and other things, Basque *aho* also means "articulation, diction" among other meanings, while Laz *p'ici* also means 'face, front side' and similarly, San Lucas Quiaviní Zapotec *ru'uh* also 'in front of' inter alia. Abzakh Adyghe *że* also means 'to grill, fry' and 'to wait' and perhaps 'knee' and 'physical deformation,' Ket *qō* also 'ice' and 'ten,' while Khalkha *ama(n)* is also used as a unit for counting persons in census or as food-consumers. Sora *tˀo:dən ~ 'tˀo:dən ~ 'tˀu:dən* may also refer to the 'jaw' (though from the source it is not clear whether the term assumes this meaning on its own or only in compounds). Haida colexifies 'mouth' with 'to feed,' and the relevant Itzaj term also denotes a tree species. Lakhota *í* also means 'to arrive at a place away from here,' and Santiago Mexquititlan Otomí also 'to love.' Abipón *n-aag-Rat* is analyzable as 'POSS.INDEF/3SG-bite-CAUS,' the association with 'bite' (as well as 'lick') is also recoverable for Nunggubuyu. Pawnee *haakaʔuʔ* also means 'drum,' while Yana *bal(la)* is also used dialectally with the meaning 'cheek' (an association evidenced by semantic shift in Indo-European, Buck 1949: 228). Kaingang *jẽnky* appears to be analyzable as 'eat-cut/smell.' Miskito colexifies 'mouth' with 'space, center,' 'road,' 'mandate,' 'message,' 'will,' and 'quarter,' Wayampi *yulu* also denotes the mouthpiece of a flute, and Wichí colexifies 'mouth' with 'hunger.' Bislama *maot* also means 'vagina' in informal use and has a further meaning related to the preparation of Kava. Kapingamarangi *ngudu* also means 'to melt' as a verb. Hawaiian colexifies 'mouth' with 'talky person,' 'neck of dress,' 'to carry on back,' 'load carried on back,' among still other meanings. Rotuman *nuju* also can refer to a 'spokesperson' and "the most easily pierced in the three 'eyes' of a coconut-shell" inter alia.

125. The Phlegm

Representation: 49%
Motivated: 51.3%
Thereof Analyzable: 12.2% Thereof Polysemous: 39.1%
Thereof by Contiguity: 6.6% Thereof by Similarity: 15.0%
Recurrent associated meanings: saliva/spittle, cold/flu, snot, cough, pus, resin, semen, throat

Eight sampled languages, Hausa, Noni, Kyaka, Ngaanyatjarra, Jarawara, Miskito, Tehuelche, and Toba colexify 'phlegm' (or 'sputum') with 'flu' or 'cold,' while the same lexical item in Embera yields the two meanings depending on gender. Fifteen languages, Khoekhoe, Ngambay, Berik, Badaga, Khalkha, Haida, Quileute, Aguaruna, Arabela, Cashinahua, Cavineña, Cubeo, Guaraní, Huambisa, Ancash Quechua, Hani, and Hawaiian, with 'saliva, spittle' (Khalkha also with 'tears'), and in Embera, again gender is used as a device to differentiate the meanings. Hausa, Ngaanyatjarra, Badaga, Basque, Itzaj, Pawnee, Tuscarora, Embera, Tehuelche, and Fijian colexify 'snot.' Kaluli and Wintu colexify 'phlegm' with 'cough,' Muna with "cough up and spit out phlegm," Toba has a term for 'phlegm' derived from 'to cough,' Bislama has *doti blong kof* 'rubbish/pus POSS cough/cold' (compare colexification of 'phlegm' and 'pus' in Badaga and Wintu), Samoan *fatu-tale* is analyzable as 'heart-cough,' and a semianalyzable term where the identifiable constituent is a verb meaning 'to cough up' is found in Kashaya. Chayahuita has *iro ipirin-so'* 'cough/flu spit.out-3SG.SUB.' Furthermore, a semianalyzable term involving a constituent meaning 'nose' is found in Yir Yoront. Pawnee *kitutkuutuʔ* is analyzable as /kitut-kuuʔat-uʔ/ 'throat-rotten-NOM,' and Lesser Antillean Creole French has the (suspiciously long) term *sistans épé ki ka sòti an né ében gòj* 'substance thick REL PROG get.out in nose or throat;' there is a semianalyzable term containing an element meaning 'neck, throat' in Highland Chontal. San Lucas Quiaviní Zapotec colexifies 'phlegm' with 'semen,' and in Burarra, *gungulol* 'phlegm' is derived from *gulol* 'semen, rotten.' Hawaiian *hākelo* also denotes 'sticky sap from trees' (there is also another term, *male*, which has other seemingly unrelated meanings such as 'to marry' and also denotes a fish species), and in Toba, 'phlegm' and 'resin' generally are colexified.

Other associations include: Hausa *kaki* colexifies 'phlegm' with 'beeswax' among other meanings, and *majina* also means 'to blow the nose' inter alia when used verbally. Ngambay *wenren* also means 'vomit,' while Kwoma *ukwa siik* contains *siik* 'fruit.' Mali *genaing* also means 'cerebrum,' and Kosarek Yale *sikna* also 'dirt, bits of dirt.' Nivkh *čjevčjevu mif* is analyzable as 'be.wet soil.' Lake Miwok *łéeʔkaṭi* also has the verbal readings of "to cough up phlegm" and for phlegm "to come up by itself in the process of clearing one's throat," Carib *kòwe* also denotes the 'round-worm,' and Rama *síri* seems to consist of the elements *sí* and *ri*, which both make reference to 'water.' Muna *ngallangalla* also means 'phloem.' In Hani, *zaoqpeiv* is 'phlegm, spittle;' both *zaoq* and *peiv* have meanings on their own that do not seem to bear a semantic relationship to either 'phlegm' or 'spittle,' the closest among the meanings of *peiv* is 'to have a sickness or disease that continues, to have

a relapse of sickness' and its function as a classifier for dewdrops. Finally, Lenakel *nouanelpwa* is analyzable as 'fruit-fat/grease.'

126. The Navel

Representation: 91%
Motivated: 28.1%
Thereof Analyzble: 8.3% Thereof Colexifying: 20.5%
Thereof by Contiguity: 17.2% Thereof by Similarity: 5.5%
Recurrent associated meanings: umbilical cord, center, placenta/afterbirth, belly/stomach, family/relatives, cause/origin

The most common lexico-semantic association for this meaning is that with 'umbilical cord.' It is realized by colexification in Buli (where the term also colexifies 'stalk of calabash' inter alia), Kwoma, Lavukaleve (where the term also means 'moonshell'), Meyah, Rotokas, Toaripi, Sahu, Sko (the relevant term is *kóengri*, for which compare *kóeng* 'tooth'), Waris, Kosarek Yale, Yir Yoront, Khalkha, Cheyenne, Ineseño Chumash, Haida (where the term also colexifies "mouth of sea urchin"), Nuuchahnulth, Quileute, Yana, San Lucas Quiaviní Zapotec, Aymara, Carib, Bislama, Hawaiian, Lenakel, Manange, and Rotuman. In Cubeo, the same root suffixed with different classifiers yields the meanings 'navel,' 'umbilical cord,' and 'liana' respectively. Miskito has *tukta awa* 'child ribbon,' suggesting that 'navel' and 'umbilical cord' are also colexified here. Furthermore, there is an analyzable term in Kashaya (*ʔohqomo* /ʔohqo-ʔimo/ 'umbilical.cord-hole'), and a semianalyzable term in Bwe Karen; in San Lucas Quiaviní Zapotec, the relevant term also may refer to a "hole in a container of liquid or a small dam that can be closed with a plug or stopper." 'Navel' is colexified with 'stomach' in Basque, and dialectally, the meaning of Kyaka *kumu* varies between 'abdomen' and 'navel' (the term also denotes a kind of tree), and complex terms betraying this association are featured in Toaripi (*ére lalave* 'belly corner'), Abzakh Adyghe (*nəbepcʔeəy*, containing *nəbe* 'belly, stomach' and *pcʔe* 'false'), and Cayapa (*ajbundyu*, presumably analyzable as /ajca-bundyu/ 'abdomen-knot'). Kwoma, Waris, Cheyenne, and Wintu colexify 'navel' with 'placenta' or 'afterbirth,' and Kyaka has *reme lyolo* 'placenta end.'

Moreover, seven sampled languages, Hausa, Kyaka (by the term *mumbi renge*, which appears to contain *mumbi* 'kind of reed' and *renge*, meaning 'source, origin, reason' alongside other things, for which compare Efik *i'buüt nsia*, analyzable as /i'buüt nsi'a/ 'head/top/cause/origin bowels'), Khalkha, Welsh, Guaraní, Hawaiian, and Malagasy extend terms for 'navel' to the abstract, namely to 'center' or like meanings (see Buck 1949: 248 for Indo-European data; Welsh in addition also colexifies 'heart'). Hausa, furthermore, also colexifies "central supporting pillar of a ceiling" and 'birthplace' inter alia, and Sedang, similarly, denotes the "center of chúa tô´u weaving pattern" inter alia, and Yay *duaˡ* also means 'center of flower.' Finally, in Khalkha, Wintu, and Hawaiian, relevant terms can also refer to the 'family' or 'relatives.'

Other associations include: Buli *siuk* also means 'street, road' inter alia. Khoekhoe *sunis* is presumably related to *suni* 'to joint, cut meat at joint,' (similarly, Wintu *naq* might

be related to *nEk* meaning 'cut off' inter alia; *naq* also means 'to pity, care for, feel sentiment, empathy towards'), and Hausa *cibi* is also the name for a small metal spoon as well as for 'umbilical hernia.' Buin *ituge* also denotes the "anal vent in fish," Burarra *jirrcha* also the tuber of the water lily *Nymphaea violacea*, and Kaluli *ho:nduf* contains *ho:n* 'water.' Kwoma *nawaba* also means 'stem of plant' and 'trunk of tree,' 'carcass,' as well as "neck of a slitdrum." Ngaanyatjarra *nyuntjirn(pa)* also means "spinning through the air very fast (of spear)," while Nunggubuyu *mala* also means 'honey' and '(clear) sky.' The Southeastern Tasmanian term *tünā* is also the name of the 'triton shell,' while Badaga colexifies 'navel' with 'umbilical rupture,' Khalkha with 'isthmus' inter alia, and Welsh with 'nave of wheel' and 'boss of shield.' Chickasaw *ittialbish ~ ittihalbisch* may contain *iti* 'mouth,' and Itzaj *tuch* also denotes the 'gizzard of birds.' Kiowa *pʻeipʻout* may contain *pʻei* 'vulva,' while Pipil *-xi:k* also means 'bell clapper' in the dialect of Santo Domingo de Guzmán, and Tuscarora *uʔnihsè·reh* also means 'star, planet.' Wintu *bo·s*, derived from *bOs* live, reside, remain, keep; stay, be in a sitting position, sit, dwell, stay,' also means 'home, camp,' and Central Yup'ik *qallaciq* is perhaps analyzable as /qalla-(u)ciq/ 'be.boiling-condition.of.possessor.with.respect.to,' with the presence of the postbase (see § 4.4.2.) not being entirely clear. Chayahuita *imëra* seems to be analyzable as /imërin-ra/ 'smell-CLASS.SMALL.THINGS,' and Kaingang *nũg-nin* as 'twist-tuber/potato.' Rama has a semianalyzable term containing an element meaning 'eye,' and Tehuelche *wet'* is also one of the variants of *wet' ~ wet ~ wet'e* 'to eat something.' For Fijian *vicovico* compare *vico*, the name of a variety of wild sugar cane. Kapingamarangi *bida* is also used with the meaning 'end' and 'piece,' and Mandarin *qi2* also means 'to ride' (reflecting different Early Middle Chinese etyma, Pulleyblank 1991: 246). Finally, Rotuman *pufa* also denotes a "shallow fire-hole for open fire, fire-place."

127. The Neck

Representation: 96%
Motivated: 39.9%
Thereof Analyzable: 2.9% Thereof Colexifying: 36.9%
Thereof by Contiguity: 30.3% Thereof by Similarity: 4.7%
Recurrent associated meanings: throat, nape, voice, collar, neck of vessel, top, behind, Adam's Apple, top, rain, swallow, head

Many sampled languages colexify 'neck' with 'throat' (see also Buck 1949: 231-232 for Indo-European data). This is the case in Efik, Berik, Burarra (also with 'larynx'), Gurindji, Ngaanyatjarra (where the relevant term denotes the front of the neck specifically), Muna, Sentani, Toaripi, Yir Yoront (where the relevant term may also refer to the "upper chest area," "mid-trunk area," and has metaphorical uses), Greek, Kildin Saami, Welsh, Highland Chontal, San Mateo del Mar Huave, Itzaj, Kiliwa, Kiowa, Wintu, Yaqui, Yuki, San Lucas Quiaviní Zapotec, Copainalá Zoque, Embera, Lengua, Ancash Quechua, Tehuelche, Tsafiki, Hawaiian (by the formally redundant term *kani-ʻāʻī* 'hard-neck'), Bwe Karen, Mandarin, White Hmong, Rotuman, and Tetun (the relevant Rotuman term also means "to squeak, to chirp" as a verb). In some of these languages, namely Burarra, Gurindji, and Muna, 'voice'

is in addition colexified. This is also the case, without colexification of 'throat,' in Katcha, Koyraboro Senni, Ngambay (which colexifies also 'melody'), Rotokas, Aymara, and Cayapa.

Abipón, Aguaruna, Arabela, Aymara, Piro, Ancash Quechua, Tehuelche, Toba, Wayampi, and Hani colexify 'neck' with 'nape' (Xicotepec de Juárez Totonac furthermore with 'nape of animal' in particular), and in Cubeo the meanings are expressed using the same root suffixed with different classifiers. This pattern is also attested in Indo-European (Buck 1949: 231-232). Haida in specific contexts colexifies 'neck' with 'head,' while Rama has a term betraying this association by a complex term (*kíngkat* /kíng-kát/ 'head/top-tree/foot'). Yuki and Hawaiian colexify 'neck' with 'Adam's Apple' (alongside 'throat,' note also that Wappo *hóch* is glossed as "front of the neck, area of the Adam's apple"). Rotokas *kopa iro* is analyzable as 'swallow/ingest belt' and colexifies 'windpipe,' 'trachea,' and 'esophagus,' while Tehuelche colexifies 'neck' and 'to swallow' (alongside other meanings) directly.

Basque, Bezhta, Haida, Tuscarora, San Lucas Quiaviní Zapotec, Bislama, Hawaiian, and Malagasy colexify 'neck' with 'collar' (and Greek with "neckline of garment" and Hawaiian also with "neck of a shirt or dress," see also Buck 1949: 232 for this association in Latin and Middle High German; it may also be present in Armenian), Buli, Muna, Abzakh Adyghe, Badaga, and Basque with 'neck of vessel' (often in particular of a bottle), and Abzakh Adyghe and Bororo with 'top' in general (note also the Rama term mentioned above and that Embera *nrɨ̃ mʉ̀sí* appears to contain *nrɨ̃* 'above'). The relevant Ngambay and Ngaanyatjarra terms can also assume the meaning 'behind.' Kyaka colexifies 'neck' with 'stem,' Kosarek Yale with "lower part of tree trunk," and Yanomámi has *ora-hitho* 'upper.extremity-petiole' (compare also the formally redundant Pipil term *kech-ku:yu* 'neck-plant/stem;' the simplex *kech* occurs only in compound); moreover, Carrier *-tsîltcen* contains *tcen* 'stick.' Finally, presumably accidentally, Kapingamarangi and Samoan colexify 'neck' with 'rain.'

Other associations include: Efik *itöñ* is inter alia also used to refer to a "connecting part," Hausa *wuya* also means 'ford' in the dialect of Sokoto, and Koyraboro Senni *jinde ~ jinda* also 'prow of boat.' Ngambay *gu* also means 'chest,' and *ngàndàra* is also an ideophone meaning "with a hoarse voice." Buin *ou* also means 'carry on shoulders,' and is the name of "a form of binding for sago thatch" inter alia, while Burarra *jawa*, as a verb, means 'to bleed, exude fluid.' Muna *dodo-ha* is analyzable as 'cut.off-LOC,' the basis for the association being that the neck is the "spot where animals are cut." Abzakh Adyghe *ləcˀ* is analyzable as /l(e)-cˀ(e)/ 'flesh-body.related,' and *pṡe* also means 'door,' 'to knead, kneading,' and 'to creep.' Basque *lepo* can also refer to the 'shoulder' and 'back.' Laz *ali* also is used with reference to any sort of constriction, and Sora *'saŋka:n* also means 'fathom.' Cashinahua *texu* is also used with the meaning 'handle of a stool.' Cavineña *erumu* also means 'landmark' and 'district' (original Spanish gloss is 'coto'), Kaingang *nunh* also means 'to pass over to another path,' and Wayampi *alɨpɨ* also means 'grape of fruit.' Fijian *domo*, as a verb, means 'to desire, lust after.' Presumably by homonymy given the shortness in terms of segments, Lenakel *rou* also means "[c]hase (in order to catch)." Mandarin *jing3* also means 'spring, well' (due to phonological collapse of erstwhile distinct etyma, Pulleyblank 1991: 159), and Lesser Antillean Creole French *kou* also 'to strike, hit' among other meanings.

128. The Nipple

Representation: 64%
Motivated: 74.2%
Thereof Analyzable: 58.2% Thereof Colexifying: 17.2%
Thereof by Contiguity: 25.7% Thereof by Similarity: 26.8%
Recurrent associated meanings: breast, point/tip/end, milk, udder, eye, head, face, suck, mouth, button, nose, seed, front

Unsurprisingly, the most common association for this meaning is that with 'breast' (and sometimes also 'teat,' an association disregarded in the following). It occurs by colexification in Dongolese Nubian (where the relevant term *ért(ı)* also means 'dirt, dirty'), Swahili, Yoruba, Lavukaleve, Northeastern Tasmanian, Basque (with optional complex terms of the major types discussed below), Khalkha, Kolyma Yukaghir, Ineseño Chumash, Nuuchahnulth, Wintu, and Sedang, which colexifies also "to pound, to beat." By far more frequent, however, are complex terms of the lexical type betraying this association. Within this class of terms, the most frequent meaning of the second element is 'tip,' 'end,' or 'point,' as in Kaluli *bó si* /bó sí ~ tí/ 'breast tip.' Terms with such structure are also attested in Buli, Rendille, Baruya, Kyaka, Ket, Biloxi (dubiously), Chickasaw, Haida, Kashaya, Kiowa, Lakhota, Lesser Antillean Creole French, Carib, Guaraní (where an additional diminutive suffix is present), Jarawara, Miskito, Wichí, Lenakel, and Rotuman; a variant of this pattern is found in Kapingamarangi, which has *mada-lili* 'end-seize.' In addition, the association is realized formally by colexification in Hawaiian, and Rotuman colexifies 'end' with 'nose,' which is why its term also bears an association with this meaning; Ngaanyatjarra colexifies 'nose' and 'nipple' directly, with secondary associations to 'beak' etc. typical for terms for 'nose.' Terms like Yir Yoront *thayn-mel* 'breast-eye' are also frequent, occurring also in Badaga (where the term denotes the 'areola' rather than the 'nipple' itself), Kiliwa (where in one variant, an additional constituent meaning 'ball' is present), Bislama, Fijian, Samoan, Takia, and Tetun, while 'nipple' and 'eye' are colexified directly in Burarra and Hawaiian (with typical other secondary associations for 'eye,' compare § 6.2.3.1.). Furthermore, Jarawara has *yohari noki/yohari noko*, where *yohari* is 'breast' and *noki/noko* colexifies 'eye' and 'seed' inter alia. 'Seed,' 'nut,' 'kernel,' and 'nipple' are colexified in Gurindji. Burarra also has the alternative term *dombu angalginy* 'developed nipple' containing *dombu* 'mudskipper' (a type of amphibious fish with protruding eyes) and *galginy* 'eye.' Due to colexification with 'eye,' there are also associations with 'face' by complex terms in Toaripi, Jarawara, Fijian, Hawaiian and Samoan, and by colexification in Ngaanyatjarra.

There are further common body-part metaphors realized by complex terms with 'breast' acting as the contiguity anchor. Copainalá Zoque, for instance, has *tzu'tzi-pocac* 'breast-head;' such terms are also featured in Kosarek Yale (where 'head' is colexifed with 'front,' compare for this Toaripi *omopa kō* 'front/face breast/udder/milk' and *kō opa* 'breast/udder/milk front/face'), Itzaj, Yaqui, Hupda, Ancash Quechua, Toba, Malagasy, and Mandarin, which also has the variant *nai3-tou2* 'milk-head.' Efik colexifies 'nipple' with 'mouth' (and the term also has other secondary associations due to the meaning 'mouth,' cf. section 124 and § 6.2.3.2.), Buli has *biisi noai* 'breast mouth/tip,' Wappo *huy-nán* 'breast-

mouth,' and Yay *paak*^{2} *nen*^{3} 'mouth breast,' and there is a semianalyzable term in Guaraní. Japanese has *chichi-kubi* 'breast-neck.' Other complex terms with 'breast' include Katcha *bibala ma thonogo* 'child GEN breast,' Mbum *nják-pâm* 'entrance/verge-breast,' Dadibi *bono si ame* 'rope two woman.breast' for 'man's breast nipple' specifically, One *nimna sola* 'breast flower,' Sora *ɔjoˑˈmeːn* /ə-ˈjoː-meː-n/ 'POSS-unripe.fruit-breast-N.SFX,' Cheyenne *he'e-vone* 'female-breast,' Kiliwa *ny+mayu=p+ʔuwn* 'POSS+breast=MP+ball/knot,' Pawnee *iit-paca* 'breast/suck-be.a.bump,' Hani *aqqul qulsiq*, derived from *aqqul* 'breast, milk, sweet' by partial reduplication and the classifier for round things *siq*, Manange *3ŋjokro 2pʰuŋ* 'breast egg,' Tetun *susun-lahat* 'breast-shrimp.net' (though note that *laho* 'mouse, rat, muscle' has a variant *laha*), and White Hmong *txiv-mis* 'fruit-breast.' There are semianalyzable terms in Guaraní and Great Andamanese, and Yanomámi has *suhe u ka* 'breast liquid opening,' with *suhe u* meaning 'milk.' This, in fact, is another natural major association for the 'nipple,' as already seen in terms from Kyaka, Hani, and Mandarin. 'Nipple' and 'milk' are colexified in Yoruba, Lavukaleve, Basque, Kolyma Yukaghir, Ineseño Chumash, Nez Perce, Nuuchahnulth, and Wintu, while Santiago Mexquititlan Otomí has *do-ˈba* 'stone-milk' and Central Yup'ik *emulek* /emuk-lek/ 'milk-one.having;' the association is also present due to colexification of 'milk' and 'breast' in analyzable terms in Kyaka, Toaripi, Kosarek Yale, Kiowa, Fijian, Lenakel, Rotuman, Samoan, Tetun, and Bislama.

Further, Khoekhoe, Yoruba, Khalkha, Nez Perce, and Wintu colexify 'nipple' with 'udder' (Khoekhoe with 'nipple of man' specifically), and the association is realized by analyzable terms due to colexification of 'udder' with 'breast' in Kyaka, Toaripi, Lesser Antillean Creole French, Toba, and Rotuman. There may be further cases of this association, namely those in which terms for 'nipple' are glossed also as 'teat' or 'teta' in Spanish; these are not taken into account here because without further evidence it is not clear whether this gloss is merely a near-synonym for 'nipple.' Badaga and Khalkha colexify 'nipple' with 'button' alongside other meanings, among them 'bullet' in Khalkha. Kolyma Yukaghir *ibiši:*, colexifying 'nipple' with 'breast' and 'milk,' is derived from a verb meaning 'to suck' (compare the Pawnee term just mentioned above). The association is present due to colexification with 'breast' in Itzaj, Lesser Antillean Creole French, and Bislama, and a semianalyzable term of this kind is found in Buin.

Other associations include: Kanuri *nzàmbòràm* contains elements meaning 'place of' and 'birth,' Kosarek Yale *selkedek* is the name for a species of raspberry, and is used with the meaning 'nipple' in dancing songs. Badaga *ma:ru* is also a unit of measurement. Greek *rṓga* also means 'grape' and *thīlḯ* also 'papilla,' while Cheyenne *he'enénestôtse* /he'e-nén-hestôtse/ is analyzable as 'female-nurse-thing.' Nuuchahnulth *ʔinma* also denotes "Nob Point, where white powder seems to run out of a breast-shaped rock," while Oneida *ohníhsyaʔ* has been extended to also denote 'baby bottle.' Santiago Mexquititlan Otomí *ts'u̲* also means 'tail.' The Wintu root *ʔEm* also bears the meaning "hold pectorally, carry something in the arms, embrace." Embera *hú*, associated with different genders, can also mean 'cavity, hollow,' and 'chest,' Piro *gikatu* also means 'stalk,' and Wayampi colexifies 'nipple' with 'stamen.' Finally, Hawaiian *ʻōmaka* inter alia also means 'building, beginning, source' and 'foreskin.'

129. The Nostril

Representation: 70%
Motivated: 78.7%
Thereof Analyzable: 76.8% Thereof Polysemous: 2.9%
Thereof by Contiguity: 57.1% Thereof by Similarity: 9.9%
Recurrent associated meanings: nose, hole/opening, mouth, burrow, cave, track, eye, house

Terms for 'nostril' are overwhelmingly complex, consisting of the respective terms for 'nose' (ignoring secondary meanings such as 'beak' etc. in the ensuing discussion) and 'hole' or 'opening,' as in Laz *çxindi xunťula* 'nose hole.' Terms with this structure are also found in Buli, Efik, Hausa, Kanuri, Khoekhoe, Mbum, Swahili, Yoruba, Anggor, Berik, Gurindji, Kwoma (where the relevant term also means "hole bored in nasal septum"), Lavukaleve, Mali, Muna, One, Sentani, Toaripi (where 'opening' is colexified with 'mouth,' compare also Takia *ŋdu-n awa-n* 'nose-3SG mouth-3SG;' terms betraying an association with 'mouth' are also featured in Lesser Antillean Creole French, Pawnee, and Malagasy), Basque, Japanese, Ket, Welsh, Kolyma Yukaghir, Biloxi (where the meaning of the other constituent next to 'nose' is more precisely "natural orifice in the human body"), Upper Chehalis, Chickasaw, Highland Chontal, Haida, Itzaj, Kashaya, Kiliwa (for the 'inner nostril;' there is the term *phi?=kw-s-?+han* 'nose=WH-INST/LONG-DN+rise.slightly' for 'outer nostril'), Kiowa, Lake Miwok, Lesser Antillean Creole French, Nez Perce, Oneida, Santiago Mexquititlan Otomí, Pawnee, Wintu, Yaqui, Bora (where segmentation is uncertain), Bororo, Carib, Guaraní, Hupda (where 'hole' is colexified with 'house,' for which compare Rama *ngú-ri* 'house-wet'), Jarawara, Miskito, Piro (where 'hole' is colexified with 'anus'), Imbabura Quechua, Toba, Tsafiki, Yanomámi, Bislama (where 'hole' is colexified with 'earth oven, cooking pit'), Fijian (where 'hole' is, as in Wintu and Sko, colexified with 'cave,' compare also Aguaruna *nuhí waa-ŋ-ĩ* 'nose cave-POSS-3POSS'), Hani, Hawaiian (where 'hole' is colexified with 'door' and other meanings, and the complex term can also refer to a "hole in pearl-shell shank"), Manange, Mandarin, White Hmong, Rotuman, Tetun, Vietnamese, and Yay. Semianalyzable terms containing an element meaning 'hole' are found in Upper Chehalis, Cayapa, and Wayampi, and one featuring a classifier for holes in Arabela. In Buli, Khoekhoe, Kwoma, Basque, and Toba, 'hole' is colexified with 'burrow,' and hence their complex terms for 'nostril' also betray an association with this meaning. Further variants of the pattern are Badaga *sivilu toḷḷe* 'breath hole.for.nosering.or.earring,' and San Mateo del Mar Huave *miwirij oxing*, containing *wirij* 'hole' and *oxing* 'point.' Similarly to the situation in Piro, in Chukchi, the relevant term may be connected etymologically to 'anus,' which may itself have born the general meaning 'hole' at an earlier stage in the language's development. Baruya has *ꞩiduta* /ꞩinna-tuta/ 'nose-track,' and a term with this structure is also attested in Kyaka. Katcha has *ɔe mɔ mbɔrɔ* 'eye GEN nose,' and a term with identical structure is found in Bislama; the pattern is also etymologically detectable in Tuscarora. Moreover, Swahili has *mwanzi wa pua* 'bamboo of nose,' Meyah *osúm efesí* 'nose interior,' Nivkh *vix kut'* 'nose fall/come.out,' Pipil (Santo Domingo de Guzmán dialect) *-bentanah-yak* '-window-nose,' Chayahuita *nitëana*, analyzable as /nitë'-ana/ 'nose-

CLASS.AROUND,' Wichí *tonhespe'*, containing *nhes* 'nose' and a locative suffix meaning 'on,' and Great Andamanese *igchōrongal'ârjâg*, containing *igchōronga* 'nose' and *jâg* 'chink, gap.'

Ngambay, Yir Yoront and Embera colexify 'nose' and 'nostril' directly (Yir Yoront also "nose-like protrusions at the fronts of animals;" there is a redundant complex term with 'hole' for 'nostril' specifically), in Cubeo, the same root yields the meanings 'nose' and 'nostril' (and "'nose' of an airplane") depending on the classifier suffixed, and semianalyzable terms where the identifiable constituent is 'nose' are found in Bezhta, Upper Chehalis, and Cavineña.

Other associations include: Sko *loelóng* appears to contain *long* 'hole, cave' and, curiously, *loe* 'ear.' Welsh *ffroen* also means 'muzzle of gun.' Central Yup'ik *curlu ~ curluq* also means 'sinus' and 'head of pike fish,' and *paciguaq*, another Central Yup'ik term for the concept, is analyzable as /pacik-(ng)uaq/ 'gill-imitation.'

130. The Pupil

Representation: 48%
Motivated: 75.4%
Thereof Analyzable: 65.7% Thereof Colexifying: 9.6%
Thereof by Contiguity: 27.4% Thereof by Similarity: 40.1%
Recurrent associated meanings: eye, black, eyeball, child/son/daughter, seed/grain, iris, small, egg, person, spot

The 'pupil' is a meaning expressed frequently by metaphor-driven complex terms (Tagliavini 1949, Brown and Witkowski 1981), with 'eye' acting as the contiguity anchor. Of these, the association with 'child' or more specifically, 'son' or 'daughter' is most frequent, due to the small reflection of oneself in the interlocutor's eye (see Urban forthcoming for discussion of the probable cultural underpinnings). For instance, Berik has *nue tan*, presumably analyzable as /nue tane/ 'eye child/niece/nephew.' Such terms are also found in Katcha, Mbum, Kaingang, Toba, and Tetun, while Basque colexifies 'pupil' with 'child, doll,' and Greek with 'girl, daugther, virgin.' Moreover, there is a semianalyzable term in Yay, and Sedang has *kón ngái*, where *kón* is 'gibbon, spider monkey,' but note *kon* 'child.' There are also other associations with human beings, namely in Muna (*ka-mie-mie-no mata* 'DIM-person-DIM-POSS eye,' Kolyma Yukaghir (*aŋd'ən-šoromə* 'eye-man'), Blackfoot (by colexification), Fijian (*yaloyalo turaga* 'reflection/image chief'), and Hawaiian (*ki'i 'ōnohi* 'image/doll eyeball;' *ki'i* also has some other meanings). Buin *tuutuu*, reduplicated from *tuu* meaning 'water' inter alia, is also an epithet for (plumb) children. Buli has *num-buli*, where *num* is 'eye' and *buli* can refer to a 'kid, young goat,' but also to the Buli language inter alia. Otherwise, 'seed' or 'grain' is common as the source concept, as in Copainalá Zoque *witʌmbuj* /witʌm-puj/ 'eye-seed/pit.' Such terms are also found in Hausa, Kyaka (where 'seed' is colexified with 'egg' and other meanings, which is also the case in Wintu; compare Yoruba *ẹyinjú* /ẹyin-ojú/ 'egg-eye'), Ancash Quechua (where the term colexifies 'iris,' as is the case in Khoekhoe, Rendille, Baruya, Wintu, and Cubeo), Guaraní (though note that *ta'ŷi* 'seed' is a diminutive of *ta'y* 'son, clot'), and Tsafiki; Biloxi has *tûtcû^{n}´su´sŭpi´* /tûtcû^{n}´su sŭpi´/ 'eye seed black.' In fact, 'black' is the major non-metaphorical association for the

meaning 'pupil.' Complex terms such as Sora *gajar-'mad-ən* 'eye-black-N.SFX' (here, the analysis is tentative; the first constituent may also be *ga'jar-* 'to turn, to reel,' 'to contain a hole'), are featured in Dadibi, San Mateo del Mar Huave, Kiowa, Nez Perce, Aguaruna, Piro, Rama, and Great Andamanese, while Upper Chehalis has a complex term involving a root meaning 'black' and a lexical affix meaning 'ridge, basket, trap' (incidentally, a term with the same lexical affix also exists based on the root for 'white'). Lenakel has *nouanhalnakan̈mr-i̱n ir-ap̈i̱n* 'eyeball-?? NMLZ-black/dark,' and furthermore, Rendille has *daáyto* /daáyi-to/ 'black-thing' and Hani *miavneev neev-nav* 'eye/eyeball RED-black.' There is a semianalyzable term where the identifiable constituent means 'black' in Khoekhoe. Another recurrent non-metaphorical association is that with 'smallness,' as betrayed in Kolyma Yukaghir *jukud-aŋd'ə* 'small-eye,' also found in Highland Chontal (where the putative constituent *galninuh*, glossed as 'little one' more precisely is borrowed from Span. *niño* and the term hence very likely a loan translation), San Mateo del Mar Huave, Cavineña, and Samoan. There are also many unique metaphor-driven conceptualizations of the 'pupil,' realized by complex terms with 'eye' acting as the contiguity anchor in the languages of the sample. Dongolese Nubian has *míssɪ-ŋ-gɛ́dɛ* 'eye-GEN-circle/rim,' Kwoma *miyi noku sobo* 'eye sago.palm pale' (*sobo* also has many other meanings), Abzakh Adyghe *nekᵊ* /ne-kᵊ(e)/ 'eye-middle,' Badaga *(kaṇ) maṇi* '(eye) gem,' Welsh *cannwyll y llygad* 'candle DET eye,' Itzaj *tz'u' ich* 'center/heart eye,' Kashaya *huʔu· ɨihta*, /huʔuy ɨihta/ 'eye bird,' Tuscarora *yučisnúhkuʔ ukáhrakę́w*, containing the roots *-čisnu̱hkw-* 'comma, speck, spot,' *-ur-* 'to cover,' and *-kah(r)-* 'eye' (for the association with 'spot,' compare colexification of 'spot,' 'fawn,' and 'pupil' in Wintu), Aguaruna *jií winchájiˀ*, containing *jií* 'eye' and *winchá* 'luminous,' Arabela *namijia nunetejojua* 'eye that.which.causes.to.sparkle,' Miskito *nakra yula* 'eye companion' (with *yula* also having other meanings), Yanomámi *mamo ishiishi* 'eye coal/soot,' and Bislama *ston blong ae* 'stone POSS eye.' There are semianalyzable terms with 'eye' being the meaning of the identifiable constituent in Efik, Khoekhoe, Carrier, Upper Chehalis, Highland Chontal, Cayapa, Bwe Karen, and Sedang.

'Pupil' is colexified with 'eyeball,' where relevant by a motivated term as discussed elsewhere in this paragraph, in Hausa, Katcha, Yoruba, Welsh, Wintu, and Guaraní; compare again Lenakel *nouanhalnakan̈mr-i̱n ir-ap̈i̱n* 'eyeball-?? NMLZ-black/dark' as well as Hawaiian has *ʻōnohi maka* 'eyeball eye.'

Other associations include: Khoekhoe *âutsiłgares*, which also means 'lens of eye,' contains *âu* 'congealed, solidified,' and Kwoma *gwadiimay* is also the 'generic term for spider.' Ngaanyatjarra *tiruny(pa)* also denotes "wild onion, onion grass." Ineseño Chumash *xutaš* is also the name of the fruit of the coffeeberry (as well as 'evening star, Venus' and "Earth mother"); similarly, Haida *hldaandaraay* contains *hldaan* 'blue huckleberry.' Central Yup'ik *takvik ~ takviun* is analyzable as /taku-vi-(u)n/ 'check.fishtrap.or.fishnet-??-device.for.' Hawaiian *kiʻi ʻōnohi* can also be used to refer to a 'beloved person.' Rotuman *rito* also denotes the "young leaves (of coconut or other palm) just as they are coming out at the top of the trunk (in the centre)."

131. The Pus

Representation: 79%
Motivated: 23.4%
Thereof Analyzable: 5.6% Thereof Colexifying: 17.8%
Thereof by Contiguity: 6.9% Thereof by Similarity: 9.5%
Recurrent associated meanings: infected/infection, rot/rotten, wound/sore, semen, dirt/dirty, boil, phlegm, resin, snot, brain, blood, water/liquid

Cheyenne, Comanche, San Lucas Quiaviní Zapotec, Ancash Quechua, Wayampi, and Yanomámi colexify, by provenience contiguity, 'pus' with 'infect, (be) infected, infection;' Yanomámi also has the complex term *niyo-niyo u-pë* 'infect-RED liquid-PL.HETEROGENOUS.' Alongside Santiago Mexquititlan Otomí, which directly colexifies 'pus' and 'wound,' there are also several languages which have complex terms for 'pus' with 'wound' or 'sore' acting as contiguity anchor. Highland Chontal has *lija ga·wi* /lijabíh la·wi/ 'lime.water.from.cooking.corn wound,' San Mateo del Mar Huave *aonts najloc ~ aonts necoy* 'excrete wound' (note also that there is a semianalyzable term with a constituent meaning 'excrement' in Sáliba and that this association is etymologically recoverable in Dongolese Nubian), Kashaya *maʔšaˈ ṭʰo·*, analyzable as /maʔša ʰṭʰo·ṭ/ 'sore rot,' and Tetun *kanek-been* 'wound-liquid.' Moreover, there is a semianalyzable term where the identifiable constituent means 'sore, wound' in One. Alternatively, Tetun also has *been-tasak* 'liquid-septic,' and an association with 'water' is also found in Sora: *'gurda:n* 'pus, juice, sap' contains *dʔa:-* 'water' and either *gʔu:r-* 'to ripen' or *gu:r-*, a variant of *go:r-* "to cut vertically soft things' (for this term, note that Mali colexifies 'pus' with "white sap from the breadfruit tree" in particular). There is a semianalyzable term with 'water' as the meaning of the identifiable constituent in Sko, and one in Abzakh Adyghe which colexifies 'pus' with 'lymph.' The association with 'rot, (something) rotten' in Kashaya is not unique: Gurindji, Ngaanyatjarra, Santiago Mexquititlan Otomí, and Tuscarora colexify these meanings, and Tetun has *raan-kroek* 'blood-rotten' alongside *raan-mutin* 'blood-white.' This, in turn, is not the only association with 'blood,' which is colexified with 'pus' in Kosarek Yale.

There are also other associations with body fluids and soft body parts, which are more frequent cross-linguistically. Baruya and Manange colexify 'pus' with 'snot' (this may also be the case in Cayapa, where the terms are however not quite identical), Ket, Lake Miwok, Carib, and Yay with 'semen' (Yay with 'pus running from the flesh' specifically), and Ket and Lake Miwok colexify 'pus' with 'brain.' Badaga and Wintu colexify 'pus' with 'phlegm.'

Furthermore, Sahu and Abzakh Adyghe colexify 'pus' with 'boil,' Hawaiian has *pala-hēhē* 'yellow-boil' (*pala* can also assume the meaning 'rotten' when speaking about taro corms inter alia), and there is a semianalyzable term containing an element meaning 'abscess' in Kaingang. Ngaanyatjarra, Rama, and Bislama colexify 'pus' with 'dirt(y)' (in Swahili, the term for 'pus' is a loanword from Arabic originally meaning 'dirt, filth'). Piro colexifies 'pus' with 'resin' and 'rainbow,' and the former association is also present in Sora.

Other associations include: Bakueri *iíjá* also denotes "a bunch of oil-palm nuts." Hausa *'diwa* is also the name of a red grass species, and dialectally assumes the meaning 'anus,' while *mugunya* is also the feminine form of *mugu* 'evil.' Muna *taghirao* 'pus from acne' also means 'soft meat inside an old coconut' (compare *taghi* 'belly, stomach'?), Ket *doˀŋ* also means 'three,' Khalkha *idege(n)* also "food, nourishment, provisions," 'kernel of nut,' and 'tannin,' and for Khalkha *øgeri* and *øgesyn* compare *ø ~ øge* 'fault, roughness, unevenness.' Sora *'tule:dən* colexifies 'pus' with 'gum,' and Pipil *te:mal* is derived from a verb meaning 'fill.' The Tuscarora root *-atkęhθr-* also yields the meaning 'leather wood.' Central Yup'ik *imaq* means 'content' and also 'bullet' and 'ocean.' Wayampi appears to colexify 'pus' with 'street' inter alia, Great Andamanese *mûn* also denotes 'marrow,' Fijian *nana* is also an 'affectionate word for mother,' while Hani *biaol* also means 'to fly.' Bwe Karen *mí* also means "be full of pus" and 'be ripe' inter alia, Sedang *hẽ* also 'saliva,' and Lesser Antillean Creole French *matie* also 'matter, material' and 'topic.'

132. The Rib

Representation: 90%
Motivated: 33.6%
Thereof Analyzable: 23.1% Thereof Polysemous: 12.0%
Thereof by Contiguity: 27.4% Thereof by Similarity: 4.5%
Recurrent associated meanings: side/flank, bone, chest/thorax/ribcage, fence

By far the most common structure in motivated terms for 'rib' in the languages of the sample is for them to be analyzable of the lexical type, with constituents meaning 'bone' and 'side' or 'flank,' as for instance in Kildin Saami *jērrh't-tāxx't* 'flank-bone' (both are common associations in Indo-European, Buck 1949: 208). Such terms are also featured in Efik, Kanuri, Koyraboro Senni, Mbum, Yoruba, Sahu, Miskito (which also has the alternative term *tnaya mina* 'side dent'), Tsafiki, Wichí, Hawaiian (where the relevant term also means 'spareribs' and 'wife,' because of the biblical motive of Eve having been created from Adam's ribs), Malagasy, Rotuman, Takia, Tetun, and Yay. There are also, alternatively, some languages with complex terms involving 'bone,' but not 'side.' These are One (*nenki amna*, with *amna* meaning 'bone' and *nenki* also occurring in *moru nenki* 'vertical thatch braces;' similarly, Ineseño Chumash colexifies 'rib' with "verticals of house frame," Anggor *ŋgerembˀ* is also glossed as "frame of mbisu" and *kwansatha* in the Piro term for 'rib' *kwansathapu* means also 'ribs of roof;' this association is paralleled in Indo-European, Buck 1949: 208), Badaga (*nenjilu* /nenju-ilu/ 'chest-bone'), Ket (*ulat* /ul-aˀd/ 'straight-bone'), Kolyma Yukaghir (*nugodˀid-amun* 'thigh-bone,' colexifying 'thigh'), Kapingamarangi (*iwi di wogowogo* 'bone ART ribcage'), Manange, which has (*ŋoŋtse*) *1krẽ 2nokrẽ* '(under.waist) hips bone,' and Samoan (*ivi 'aso'aso* 'bone variety.of.yam'). Semianalyzable terms with 'bone' as the identifiable constituent are attested in Efik, Sora, the Cuisnahuat dialect of Pipil, Copainalá Zoque, and Guaraní.

Conversely, there are also languages where the relevant terms betray an association with 'side,' but not with 'bone.' Kyaka, Ngaanyatjarra, Sko, Basque, Greek, Khalkha, Nivkh, Bororo, Bislama (by a rare term), and Hani directly colexify 'rib' with 'side (of

body)' (Basque also with 'hulk' and 'point of view, viewpoint' and Khalkha also with 'wing' and "spur of a mountain between two valleys"), Chukchi has *ɣəto-lqəl* 'side-material.for,' Piro *kwansatha-pu* 'side-shape.of.bean.or.banana,' and there is a semianalyzable term in Gurindji.

Moreover, Buli, Ngambay, Buin, Rotokas, Kiliwa, Xicotepec de Juárez Totonac, Bororo (by a semianalyzable term containing the constituent 'tree' and colexifying 'trunk'), Fijian, and Mandarin colexify 'rib' with 'chest' and/or 'thorax, ribcage' (Ngambay also with 'basket' and 'measles'), while in Cubeo, the same root yields both meanings depending on the suffixed classifier, Bora has *mɨ̂jowa*, perhaps analyzable as /mɨ̂jco-gwa/ 'chest-SCM.2d.straight,' and the association is also present in the Badaga and Kapingamarangi terms mentioned above. Finally, Lavukaleve colexifies 'rib' with 'fence,' and analogously, Muna has *karakara*, with the apparent reduplication base *kara* meaning 'yard, yard fence.'

Other associations include: Hausa *hak'ark'ari* also means 'pneumonia,' Kaluli *sidif ~ tidif* appears to contain *sí ~ tí* 'tip,' and Meyah *osrój* also means 'fiancée, boyfriend, girlfriend.' Yir Yoront *pawrrmel* contains *mel* 'eye,' while one constituent of the Abzakh Adyghe term *cağe* is *ce* 'fiber, blade.' Khalkha *xabirɣa(n)*, apparently derived from *xabir- ~ xabira-* 'to whet, grind, rub, touch lightly in passing, for animals to stand close to one another,' also means 'wing,' *xabisu(n)*, which seems to be derived from *xabi* 'vicinity, neighborhood,' also means 'womb, uterus,' and Welsh *asen* also 'she-ass.' Central Yup'ik (Yukon, Lake Iliamna and Nunivak Island dialects) *inarun* is analyzable as /inar-te-(u)n/ 'lying.down-act.on.one.so.as.to.cause.it-device.for.' Arabela *riuquiocua ~ riquiocua* is analyzable as /riuquionu-cua/ 'bend.something.to.form.a.receptable-CLASS.MOULD.' Cayapa *vi'chi* might consist of *vi* 'chaquira' and *chi* 'tree, wood,' and Rama *palkát* contains *kát* 'tree, stick.' Fijian *sarisari* also denotes the "ribs" of a boat, while *waqawaqa*, denoting both 'ribcage' and 'rib,' is reduplicated from *waqa*, meaning 'box, container' inter alia. Bwe Karen *-we* is glossed inter alia as 'shore, bank, waterside' in the Bwe Karen-English section of the consulted source, while Lenakel *nakau* also means 'midrib of coconut' and 'side of mountain.'

133. The Saliva

Representation: 82%
Motivated: 34.3%
Thereof Analyzable: 18.0% Thereof Colexifying: 16.4%
Thereof by Contiguity: 11.3% Thereof by Similarity: 16.2%
Recurrent associated meanings: water/liquid/juice, phlegm/sputum, mouth, foam, spit, synovial, soup

Terms for 'saliva' ('spittle,' 'drool,' 'slaver') of the analyzable lexical type in which one constituent means 'mouth' (possibly with associated extensions, cf section 124 and § 6.2.3.2.) and the other 'water,' 'juice,' or 'liquid' more generally are relatively frequent. For instance, Yuki has *nan-uk* 'mouth-water,' and such terms are further found in Koyraboro Senni, Ngambay, Kyaka, Abzakh Adyghe, Upper Chehalis, Chickasaw, Itzaj, Kiliwa, Wappo, Jarawara, Maxakalí, Miskito, and Lenakel. San Mateo del Mar Huave has *aonts ombeayaran*

'excrete mouth,' and semianalyzable words with the identifiable constituent meaning 'mouth' are attested in Nunggubuyu and Cashinahua. Furthermore, Kosarek Yale *sulu-mak* appears to be analyzable as 'strong-water,' Chayahuita has *iro-i'* 'cough/flu-CLASS.LIQUID,' Tsafiki *pi'pí* appears to be reduplicated from *pi* 'water, liquid, juice,' Yanomámi has *kahu u* 'cover liquid,' Hawaiian *wale wai* 'slime/phlegm water,' and Bwe Karen *bəyà-chí* 'person-water.' In Khoekhoe, a term for 'saliva of animals, dribble' dialectally also has the meaning 'water,' and likewise, one for 'foam, saliva' means 'water' in the Northern dialect of Ngaanyatjarra. There is a derived term in Great Andamanese, and semianalyzable terms are found in Kosarek Yale, Ket, Bororo, Piro, Rama (the relevant constituent meaning 'wet' rather than 'water'), and Tetun ('juice' is colexified with 'water' or 'liquid' in a few of the above mentioned languages, and a semianalyzable term with a constituent meaning 'juice' specifically is found in Wayampi). In Itzaj, *k'a'* in *k'a' chi'* 'saliva' colexifies 'liquid' with 'juice' and 'soup,' while Khalkha and *silysy(n)* seems to be derived from *sily(n)* 'soup, buillon, broth' (Khalkha has another term, *nilbusu(n) ~ nilmusu(n)*, which also can refer to 'tears').

Otherwise, two languages of the sample, Sahu and Ineseño Chumash, colexify nominal 'saliva' with verbal 'to spit,' while in Khalkha and Fijian, there are terms for 'saliva' standing in a derivational relationship to a verb meaning 'to spit;' further, in Koyraboro Senni, Biloxi, Nuuchahnulth, and Tehuelche, there is an apparent relationship of this kind, but the precise mechanism of word formation cannot be identified (anymore). Seventeen sampled languages, Khoekhoe, Ngambay, Berik, Badaga, Khalkha, Haida, Quileute, Aguaruna, Arabela, Cashinahua, Cavineña, Cubeo, Guaraní, Huambisa, Ancash Quechua, Hani (the relevant term *zaoqpeiv* contains *peiv*, meaning inter alia 'to have a continued sickness, to have a relapse of sickness' and also functioning as a classifier for dewdrops; *zaoq* also has other apparently unrelated meanings), and Hawaiian, colexify 'saliva' and 'phlegm,' and in Embera, the same root associated with different genders yields both relevant meanings. Efik and Khoekhoe colexify 'saliva' with 'synovia.' Furthermore, there is a pattern of colexification, namely that with 'foam,' that is particularly frequent in Oceania. It is attested in Buin (which also colexifies 'lungs'), Kwoma, Lavukaleve, Ngaanyatjarra (which colexifies also 'soap powder'), Tasmanian (Western and Southeastern), Badaga, Lesser Antillean Creole French, Tsafiki, and Bislama.

Other associations include: Efik *u'döt*, derived from *dör'ö* 'be viscous, glutinous,' is a general term for "any viscid animal secretion," while Hausa colexifies 'saliva' with 'day, today.' Koyraboro Senni *hatta* also means to "gather spittle into throat before spitting" as well as 'to clear one's throat' and 'to miss target.' Ngambay colexifies 'saliva' with 'vomit,' and Yoruba *itọ́* also means 'small creek.' Nunggubuyu *nᵍa:l* is a term for 'slimy substance' generally, which can for instance also refer to decaying jellyfish on the beach, Rotokas *vevega* also means 'slime' and 'semen,' Sentani *ki* also 'deed, act' and 'infant,' and Yir Yoront colexifies "vapour spouted by sea-mammal, spray spouted by sea-mammal." Badaga *ecca ~ eccalu ~ enjala* also means "impurity from contact with mouth" and 'leftover food,' and Basque *listu* also 'to fray, unravel.' The Blackfoot term *sóópoyooyihkaan* contains *soopoyóóyihkaa* 'to drool,' Kiowa *t̂outk'yʜdl* contains *k'yʜdl* 'to be wet,' and Embera *íɗuɓá* contains *í* 'lip' and *ɗu* 'deep.' Guaraní *tendy* also means 'flame, light,' the Huambisa term

saawin contains *saawi* 'clear, transparent,' while Kaingang *jã-ra* appears to be analyzable as 'tooth-jaw.' Lengua *emanang* might contain *ema* 'blood,' and Great Andamanese *âkàraij* might contain *raij* 'to shoot.' Hawaiian *ʻae* also means "sap wrung from seaweeds or leaves of plants such as taro" inter alia, and *hāʻae* is also the name of "a variety of sweet potato from which bear [sic] was made." Finally, Sedang *hẽ* also means 'pus.'

134. The Scar

Representation: 78%
Motivated: 33.2%
Thereof Analyzable: 12.3% Colexifying: 21.7%
Thereof by Contiguity: 14.7% Thereof by Similarity: 9.6%
Recurrent associated meanings: wound/sore, mark, cut, trace, sign, heal, footprint, scab, welt, birthmark, spot, place, lump/lumped, depresssion, scale, stain, shadow, old

Koyraboro Senni, Yoruba, Ngaanyatjarra, Nunggubuyu, Yir Yoront, Piro, Tehuelche, and Manange colexify, by provenience contiguity, 'scar' with 'wound' or 'sore' (Yir Yoront also with 'bite of insect'). Alternatively, Sko has *píúe-lúe*, presumably analyzable as 'wound-chop/cut.small.things,' Kosarek Yale *meneng kiklek* 'wound cautious,' Basque *zauri-marka* 'wound-mark,' Japanese *kizu-ato* 'wound-trace' (this is quite a frequent colexification cross-linguistically: 'scar' and 'trace' are colexified in Efik, Yoruba, Muna, Abzakh Adyghe, Toba, and Hawaiian, and Lenakel has a semianalyzable term), Kildin Saami *avv'-sajj* 'wound-place,' Hupda *hɔm bɨ̆g* 'wound old' (compare also Wayampi *(l-)ɛna-ɛ* '(DEPENDANCE-)place-PAST'), Bislama *mak blong soa* 'mark/spot POSS sore' and *trak blong soa* 'track POSS sore,' and Tetun, like Kildin Saami, *kanek-fatin* 'wound-place.' Semianalyzable terms with 'wound' or 'sore' are found in Ngambay, Berik, Kaluli, One, Piro, Sáliba, and Hani, in which latter there are other semianalyzable terms involving a constituent meaning 'be hurt, ache, be ill.' There are also a number of sampled languages in which 'scar' bears an association with the meaning 'to cut, cut.' This is found by colexification in Koyraboro Senni (Gao dialect), Hawaiian, and Samoan, Sko has, as mentioned above, *píúe-lúe*, 'wound-chop/cut.small.things,' Itzaj has *kuuch xot'al tu'pal* 'place cut/chopped extinguished,' Santiago Mexquititlan Otomí a derived term, Aguaruna *tsupí-k-mau wakaní* 'cut-ASP-NMLZ shadow' (note also the colexification of 'scar' and 'shadow,' alongside 'companion, spouse, friend' in Cashinahua), and Kapingamarangi *di lohongo me ne wele/tuu* 'DET position thing PERF burned/cut.' Baruya, Nuuchahnulth, and Bwe Karen colexify 'scar' with 'heal(ed),' Ket has *binaʁols* containing *bin* 'self' and *qol* 'heal,' and Fijian *i macamaca* 'DERIV dry/healed.' Buli, Efik, Hausa, Yoruba, Yir Yoront, Basque, Greek, Khalkha, Toba, and Lesser Antillean Creole French colexify 'scar' with 'mark' (Yoruba also with 'impression;' similarly, Samoan has *mā-ʻila* 'COMPLETE-mark'), Yoruba, Greek, Highland Chontal, Lesser Antillean Creole French, and Yanomámi with 'sign,' Buli, Efik, Guaraní (among other meanings), and Toba colexify 'scar' with 'footprint' (furthermore, there is a semianalyzable term with a constituent with this meaning in Lenakel), Hausa and Basque with 'spot,' Tuscarora, Aguaruna and Hawaiian with 'scab' (Aguaruna also with 'sterile,' and Hawaiian also with 'button, badge' and 'blunt, dull' as well as 'fish'), Upper Chehalis and Guaraní with 'lump(ed),'

Chayahuita and Hawaiian with 'depression,' Muna and Tuscarora with 'scale,' Khoekhoe, Noni, and Greek with 'birthmark,' Yoruba and Hawaiian with 'stain,' and Muna, Nunggubuyu, and Hawaiian with 'welt.' Finally Yanomámi has a semianalyzable term for 'scar' with the identifiable constituent meaning 'skin,' also meaning 'cover' more generally.

Other associations include: Efik colexifies 'scar' inter alia with 'seam,' 'crack,' 'flaw,' and 'chink,' and *inia* is also the name of "a disease covering the body with large sores." Hausa *adabali* also denotes a "patch riveted on to a damaged sword," and *kufai* also an "old site of house, compound, or even a town" inter alia. *Tabo*, another Hausa term, is also used to refer to a favor not paid back and an old fault that can be brought up against someone. Burarra *munjakarn* contains *munjak* "hard ground on the floodplains which have dried out," and Kyaka *mumbwua* colexifies 'scar' with 'callus.' Muna *bhili* also denotes "hanging roots, supporting roots," and *ghana* also means 'to not finish, not use up.' *Pinda* is another Muna term with apparently unrelated additional meanings, among them "traditional plate made of bone." Relevant Tasmanian terms in all varieties with the exception of the Northern (for which data are lacking) also mean 'wart,' 'wrinkle,' and 'tail,' and in Yir Yoront there is a semianalyzable term where the identifiable constituent means 'mouth.' Basque *orbain* also denotes a 'pock.' Greek *sīmádi* also means 'target,' as well as 'omen, sign.' Sora *dul'dulən* appears to be analyzable as /dʊl-dʊl-ən/ 'RED-finish/bite-N.SFX,' and *gag'garən ~ gal'galən* as /gag-gar-ən/ 'RED-pierce/boar.a.hole-N.SFX;' this term also denotes the "pits of small-pox" specifically. *'Tar-pu:-n* is analyzable as 'white-stab-N.SFX,' and *pənugo:n*, yet another Sora term, is related to *'pugo·-* "to be coloured, to be scarred, to be spotted, to be overcast with dust." Blackfoot *isttsikakkssin* is derived from *isttsikakkí-* 'to form shiny scar tissue,' which in turn seems to contain *isttsi* 'pain, ache, hurt' and perhaps *wakkii* 'to heal.' Wintu *tʰaw* also means "grief, mourn, sad." Central Yup'ik *qelengllak* also means 'wrinkle' and 'kink,' and *iiraq* also 'parotid gland,' 'tonsil,' and 'side of neck.' Copainalá Zoque colexifies 'scar' with 'crab species,' whereas Cubeo *toa-churi* is analyzable as 'fire-CLASS.SCAR.OR.WOUND-LIKE.OBJECT.' Piro *ptšo̲* also means 'dirtiness.' Yanomámi *kano si* also means 'rubbish' and "abundant fruits which one collects several times from one place," and *tusi ~ tusitusi* also means "covered anew with vegetation." Bwe Karen *gəθrótha* contains *tha* 'up, rising.' Figuratively, Hawaiian *'ālina* also means 'low, disgraced, dishonored,' and *lina* also means 'soft' and 'sticky' (there are also other apparently unrelated meanings). Malagasy *hòlatra* also means 'mushroom,' Rotuman *pạtu* also "to hit (a person's head) with the back of the fingernail," while Yay *pan*^{4} *piaw*^{3} is analyzable as 'become to.sear.' Lesser Antillean Creole French *mak* also means 'stamp' and 'mole' inter alia.

135. The Skin

Representation: 97%
Motivated: 73.4%
Thereof Analyzable: 3.1%
Thereof Colexifying: 70.0%
Thereof by Contiguity: 20.0%
Thereof by Similarity: 26.0%

Recurrent associated meanings: leather/hide, bark, rind/peel, shell, surface/covering, husk/chaff, scale, body, crust, clothing/cloth, complexion, pod, fingernail, scab, flesh, race, kin/family, rubber tire

Many sampled languages colexify 'skin' with other types of outer covering of living things, often with colexification of more than one additional sense. Many languages colexify 'skin' with 'bark' (and there are many other languages in which 'bark' is expressed by a complex term on the basis of 'skin,' compare section 3; this association is only weakly attested in Indo-European, namely in Germanic, according to Buck 1949: 201). The relevant languages are Bakueri, Efik, Yoruba, Anggor, Buin, Burarra, Gurindji, Kwoma, Kyaka, Lavukaleve (colexifying 'bark of coconut tree' specifically), Mali, Muna, Ngaanyatjarra, Nunggubuyu, Sahu, Sko, Sentani (alongside a particular palmtree in its entirety), Southeastern and Western Tasmanian, Toaripi, Kosarek Yale, Basque, Bezhta, Biloxi, Chickasaw, Ineseño Chumash, Comanche, Haida, Pipil, Xicotepec de Juárez Totonac (colexifying 'skin of a person's leg' specifically), Yaqui, Copainalá Zoque, Abipón, Bororo, Carib, Cashinahua, Cayapa, Guaraní, Hupda, Jarawara, Kaingang, Maxakalí, Piro, Ancash Quechua, Tsafiki, Wayampi, Yanomámi, Bislama, Fijian, Hawaiian, Bwe Karen, Lenakel, Malagasy, White Hmong, Rotuman, and Samoan. Moreover, Cahuilla colexifies 'bark' with 'skin of animals.'

Bakueri, Yoruba, Baruya, Kyaka, Mali, Muna, Ngaanyatjarra, Basque, Khalkha, Nivkh, Welsh, Chickasaw, Haida, Lake Miwok, Lakhota, Lesser Antillean Creole French, Oneida, Pawnee, Pipil, Wintu, Central Yup'ik, San Lucas Quiaviní Zapotec, Bororo, Embera, Guaraní, Jarawara, Kaingang, Rama, Yanomámi, Bislama, Fijian, Hawaiian, Rotuman, and Tetun colexify 'skin' with 'rind' and/or 'peel,' and Efik, Yoruba, Muna, Nivkh, Cahuilla, Chickasaw, Ineseño Chumash, Lake Miwok, Lakhota, Pawnee, Pipil, Xicotepec de Juárez Totonac, Yuki, Copainalá Zoque, Bororo, Cashinahua, Cayapa, Embera, Guaraní, Hupda, Kaingang, Maxakalí, Ancash Quechua, Sáliba, Tsafiki, Yanomámi, and Bwe Karen colexify 'skin' with 'shell' of various objects, e.g. a nut or crustacean. Bakueri, Efik, Yoruba, Baruya, Muna, Rotokas, Sahu, Khalkha, Nivkh, Embera, and Bislama colexify 'skin' with 'husk' and/or 'chaff' (see also Buck 1949: 201 for this association in diachrony in Indo-European), Efik, Nivkh, Biloxi, Lake Miwok, Wintu, Guaraní, Jarawara, and Bwe Karen with 'scale' (see also Buck 1949: 201 for this association in diachrony in Celtic), Kyaka, Muna, Basque, and Rotuman with 'crust,' Efik, Yoruba, and Nivkh with 'pod,' Efik and Central Yup'ik with 'scab,' Anggor, Sko, Kiowa, and Yanomámi with 'clothing' or 'cloth,' and Highland Chontal and Aymara with 'flesh' (Highland Chontal also with 'pulp').

Still more generally, terms for 'skin' are extended to 'surface' or 'covering' in general in Kwoma, Sko, Abzakh Adyghe, Basque, Nivkh, Comanche, Haida (also with 'surface of waves'), Lake Miwok, Lakhota (colexifying also 'envelope' and 'wrapping' specifically'), Nuuchahnulth, Cubeo, Guaraní, Jarawara, Lengua, Maxakalí, Piro, Wayampi, Yanomámi, Bislama, Hawaiian, Kapingamarangi, Bwe Karen, and Lenakel (and 'cover' is the ultimate source of many Indo-European terms for 'skin,' Buck 1949: 200-201).

Noni, Baruya, Buin, Kwoma, Mali, Aymara, and Toba colexify 'skin' with 'body,' while Dadibi *tigi wali* is analyzable as 'body wind.' Ineseño Chumash, Cubeo, and Hawaiian colexify 'skin' with 'complexion' (as did Ancient Greek, Buck 1949: 200), and Ngaanyatjarra

and Khalkha with 'race.' By metaphorical extension, Miskito and Samoan colexify 'skin' with 'kin, family.' Oneida colexifies 'skin' with 'rubber' (by a term probably containing an element meaning 'piece of cloth, rag'), and similarly, San Lucas Quiaviní Zapotec *gui'ihdy ~ gyi'ihdy* also denotes 'plastic,' while Rotuman and Samoan colexify 'skin' with '(rubber) tire'

Another major association is, by provenience contiguity, that with 'leather' and/or 'hide' and sometimes also 'pelt' (see also Buck 1949: 200 for ample Indo-European evidence). This is found exclusively by colexification in the languages of the sample and is attested in Buli, Efik, Khoekhoe, Koyraboro Senni, Noni, Dongolese Nubian, Swahili, Yoruba, Kyaka, Muna, Sentani, Abzakh Adyghe, Badaga, Basque, Chukchi, Greek, Khalkha, Sora, Blackfoot, Upper Chehalis, Haida, San Mateo del Mar Huave, Kiowa, Lesser Antillean Creole French, Nuuchahnulth, Santiago Mexquititlan Otomí (by the term *xifni*, analyzable as /xi-fani/ 'body.hair-horse'), Pawnee, Pipil, Quileute, Xicotepec de Juárez Totonac, Tuscarora, Wappo, Yana, Yaqui, Central Yup'ik, San Lucas Quiaviní Zapotec, Copainalá Zoque, Abipón, Aguaruna, Arabela, Bora, Bororo, Carib, Cashinahua, Cavineña, Chayahuita, Embera, Guaraní, Huambisa, Jarawara, Kaingang, Lengua, Miskito, Piro, Ancash and Imbabura Quechua, Sáliba, Tehuelche (colexifying 'ostrich leather' more specifically), Toba, Tsafiki, Yanomámi, Hawaiian, Bwe Karen, Malagasy, Manange, White Hmong, Rotuman, Samoan, Sedang, and Tetun.

Other associations include: Buli *gbain* also means 'drumhead' and *gbang* also means 'paper' and denotes also several objects made of paper; it can also refer to 'gambling.' Efik *ik'pa* inter alia also means 'whip,' 'stripe,' and 'surface,' e.g. of water lying still. Hausa *fata* also means 'to wish well, well-wishing' and "blow the nose and project mucus into or upon something." Ngambay colexifies 'skin' and 'shoe,' Burarra *-maliyarra* 'skin' and "wrapper, outer case" (this term may be related to *mala* 'clan'), Muna *kuli* also means to "have a simple meal without side dish," Ngaanyatjarra *miri* also to "get a shock, be shocked" and *pangki* (Northern dialect) also 'orange.' One *tapi* also means 'thin,' and Meyah *ofos* is also used with the meaning 'mountain top.' Waris *tóvól ~ nihtóv* also means 'blood,' Abzakh Adyghe *ṡ°e* also 'appearance, color' and *ṡḥe* inter alia also 'head, top,' Badaga *to:lu*, similarly to One, also 'skinny,' Basque *azal* also 'sheet,' 'case,' as well as 'rascal,' Nivkh *hatx* also 'envelope,' 'film,' and 'fur,' and Biloxi *ahí´ ~ ahe´ ~ ahĕ´~ he* also means 'fingernails, toenails' (Rama also colexifies 'nail'), 'horn,' and 'hoof.' Kiowa colexifies 'membrane' and 'cloth, mat,' Wintu *ċop* also means 'acorn' and *la·* also 'tendon' and 'string, stretch.' Copainalá Zoque *naca* is also used with the meaning 'sole.' Hupda *b'ɔ́k* also means 'dish, plate, food,' and Maxakalí *xax* also to "seek, hunt, long for." Miskito *taya* also means 'feather,' Piro *mta* also 'mat,' and Rama *uk* also 'coat' and 'nail.' Toba *l'oc* may colexify 'skin, body' with 'cloud,' though *l-* in the body-part terms is a 3rd person singular possessive prefix, and it is unclear whether the similarity with *l'oc* 'cloud,' with the consonantal onset apparently belonging to the root, is merely accidental. Wayampi *pi* also means 'to finish' inter alia. Bwe Karen (-)*kó* can also mean "to swell as the result of the presence of pus or fluid under the skin" and 'mountain,' and *phe* also "to scratch, claw, maul." Hawaiian *'ili* also means 'area, land section' and *'alu'alu*, formally reduplicated from *'alu* meaning inter alia 'sag, flabby' and 'depression, gully,' has many meanings, among them 'foetus,' while Lenakel

nɨvig- may also refer to the 'cover of a book.' Samoan *pa'u* also means 'belt of machine' and 'foreskin,' and White Hmong *tawv* also means 'hard.'

136. The Snot

Representation: 75%
Motivated: 37.4%
Thereof Analyzable: 15.8% Thereof Colexifying: 21.6%
Thereof by Contiguity: 15.0% Thereof by Similarity: 14.3%
Recurrent associated meanings: nose, cold/flu, phlegm/mucus, blow nose, runny nose, water, pus, faeces

Two sampled languages, San Mateo del Mar Huave and Kashaya, have complex terms of the lexical type for this meaning with the constituents meaning 'nose' and 'excrete, faeces.' Furthermore, there are semianalyzable terms where the identifiable constituent is 'faeces' in Cayapa and Lenakel. The Cayapa term just mentioned colexifies 'snot' with 'flu' (its shape is *quijpe*, compare also *quiipe* 'pus'?), and this or colexification with 'cold' or 'have a cold' is also found in Hausa (dialectally), Buin, Ngaanyatjarra, Yir Yoront (by a semianalyzable term also containing 'nose'), Wintu, Arabela, Bora, Embera, Hupda, and Miskito. Sahu has *ma si'dangutu* 'POSS have.a.cold,' and Chayahuita *iro nitën quëran pipirinso'*, containing *iro* 'cough, flu,' *nitë'* 'nose,' and *pipirin* 'to come out.' There is a semianalyzable term containing an element meaning 'flu' in Piro. In four sampled languages, Abzakh Adyghe, Itzaj (where the term means 'clear mucus' specifically), Kiowa, and Tetun, 'water' or 'liquid' rather than 'faeces' is the meaning of the second constituent, for instance Kiowa *mą̄'-t'ǫų* 'nose-water' (a similar situation is etymologically recoverable in Tuscarora). A semianalyzable term with 'water' as the identifiable constituent is found in Mbum. Alongside the term with this structure which is used for 'watery snot,' Abzakh Adyghe also has *pe-šən* 'nose-pus' for 'slimy snot,' and 'snot' and 'pus' are colexified in Baruya and Manange. 'Snot' and 'phlegm, mucus' are colexified in Hausa, Ngaanyatjarra, Badaga, Basque (also with 'gum, resin,' and 'wick'), Itzaj, Pawnee, Tuscarora, Embera, Tehuelche, and Fijian.

Upper Chehalis has *s-ɬáp=qs* 'CONTINUATIVE-thick(of liquid)=nose/point,' Kiliwa *phi?=chiilq* 'nose=boil/eruption/pustule,' Bororo *eno bori* 'nose wax,' Carib *enata aikulu*, containing *enata* 'nose' and *aiku* 'juice,' Embera *kŭ-mór* 'nose-INTERNAL.SUBSTANCES.OR.ORGANS.OF.BODY,' Rama *táik síri* 'nose slime,' Yanomámi *hushihushihi* (compare *hushi* 'nose'), Bislama *bata nus* 'butter nose,' and Samoan *isu-pē* 'nose-dead.' Japanese has a term colexifying 'nose' and 'snot' directly, as does Bislama, which also has the optional redundant term *doti blong nus* 'rubbish/pus POSS nose.' Furthermore, there are semianalyzable terms with the identifiable meaning 'nose' in Highland Chontal, San Mateo del Mar Huave, Kashaya, Kiliwa, Pipil, and Rotuman. Finally, Hausa, Khoekhoe, Koyraboro Senni, Sora, Wintu, and Kapingamarangi have terms that are either ambiguous between nominal 'snot' and verbal 'to blow the nose' (Hausa, Koyraboro Senni, Wintu, Kapingamarangi), or derived from verbs with that meaning (Khoekhoe, Sora). Similarly,

the Acoma term is derived from a verb meaning 'to have a runny nose,' and 'snot' and 'runny nose' are colexified directly in Waris, Cheyenne, Rotuman, and Tetun.

Other associations include: Efik *mk'pö* is also a variant form of *ñk'pö* 'thing, utensil, event, cause,' Hausa *majina* also means 'face,' Buin *kina* colexifies 'coconut sprout' and 'kina' (the currency of Papua New Guinea), Burarra *an-gulol* is analyzable as 'CLASS.MASCULINE-semen/rotten,' and Sko *lóeri* might contain *loe* 'ear.' Bezhta *xida* also means 'dew,' and Japanese *hana*, with different prosodic structure, also means 'flower.' Carrier *nêninthasţłes* contains *ni* 'nostril' and *hwoţłes* 'mud.' Chickasaw colexifies 'snot' with "growth on a turkey's beak," while Pawnee *piruus* also means 'be crooked, bent.' Aguaruna *búshuk(u)* is also the name of a species of edible mushroom, while Cubeo *cõenó* also denotes the concept 'tar,' Ancash Quechua *puqru* also means 'abscess,' and Wayampi *amɨ* also 'deceased.' Hani *aqbeil beilgaol* 'liquid nasal discharge' contains *gaol*, meaning 'clean water' inter alia, and *aqbeil beilniul* 'dense or solid nasal discharge' contains *niul*, meaning 'green' inter alia. Bwe Karen *nɛkhə'ɪ* might contain *khə̀'ɪ* "a variety of edible wild fern" plus a prefix for body parts. Mandarin *ti4* also means 'to cry,' 'tear,' or 'dripping' (reflecting the same Early Middle Chinese etymon, Pulleyblank 1991: 305).

137. The Semen

Representation: 45%
Motivated: 46.7%
Thereof Analyzable: 26.8% Thereof Colexifying: 21.5%
Thereof by Contiguity: 9.8% Thereof by Similarity: 18.4%
Recurrent associated meanings: water/juice, seed, penis, pus, child, candle, egg, testicle, white, brain, resin, milk, urine

Quite frequently, terms for 'semen' are complex of the lexical type, with one constituent meaning 'penis.' As for the second element, 'water,' 'liquid,' and/or 'juice' is most common, as in Kosarek Yale *kede mak* 'penis water/juice.' Such terms are also found in Kyaka, Muna, Cayapa, Yanomámi (where 'penis' is colexified with 'seed'), and Tetun. Meyah and Hawaiian directly colexify 'water, juice' (Hawaiian also "liquids discharged from the body" generally and Meyah also 'river,' compare section 47 for this pattern) with 'semen,' Mbum has *mbìì gûn* 'water child,' Rotokas *ruve ovi* 'slimy CLASS.LIQUID,' which also means 'slime,' Guaraní *kuimba'e rykue* /kuimba'e tykue/ 'man juice,' and Bislama has *waet-wota* 'white-water' (for the association with 'white,' compare also Kiliwa *ʔ-l+sap* 'DN-ILL+white' and Hawaiian *keakea*, with the reduplication base *kea* 'white'). Nez Perce has *simqé-heqs* /símqe-heqes/ 'penis-pus,' and 'semen' is colexified with 'pus' in Ket, Lake Miwok (which also has a complex redundant term), Carib, and Yay. In Ket and Lake Miwok, 'brain' is furthermore colexified. Toaripi has *fe-oro* 'penis-lime,' and semianalyzable terms where the identifiable constituent is 'penis' are found in Kaluli, One, and Yir Yoront.

Khoekhoe colexifies 'semen' with 'urine' by an archaic term. Similarly, Tuscarora has *uʔnhęhsú·kriʔ*, analyzable as /u-ʔnhęhs-ukr-iʔ/ 'NOUN.PREFIX-egg/testicle-rubbish-NOUN.SUFFIX,' with the stem *-ʔnhęhsukr-* also yielding *uʔnhęhsú·kreh* "foul or disgusting urine, a slovenly or slatternly person so filthy as to emit an odor of urine." Associations

with 'egg' are also found in other languages: Rama has *yáat aríra* 'testicle/egg string/fold' ('testicle' and 'semen' are colexified in Guaraní), and Kiowa directly colexifies 'egg' with 'semen,' as well as with 'child' (there is also a redundant complex term on the basis of this term with the additional constituent meaning 'white'). This latter association with 'child' is also found in Wintu, Wayampi, and Hawaiian, and, as already noted above, Mbum has *mbìì gûn* 'water child;' compare also Guaraní *ta'ỹi* /ta'y-i/ 'son/clot-DIM.' Miskito and Hawaiian colexify 'semen' with 'sap, resin.' In fact *wai*, the relevant Hawaiian term has very broad reference, including 'liquid' generally, but also 'honey' and any liquid discharged from the body inter alia.

Dongolese Nubian, Basque, Greek, Welsh, Wintu, Aymara, and Guaraní colexify 'semen' with 'seed' (the Basque term also means 'breed' and has a verbal reading 'to raise, grow' inter alia, and the Dongolese Nubian term contains an element meaning 'to sow'), and San Lucas Quiaviní Zapotec and Bislama with 'milk.' Cubeo (by a term whose root suffixed with different classifiers yields 'gasoline lamp,' 'lantern,' and 'battery'), Macaguán (by the term *pitiócha*, analyzable as /pe-itiót-ja/ '3SG.POSS-enlighten-NMLZ'), and Sáliba colexify 'semen' with 'candle.'

Other associations include: Hausa has *zuwan kai* for both 'semen' and 'orgasm.' *Zuwa* is glossed as 'coming,' and *kai* has extremely many meanings, among them 'to come, arrive,' and 'head, top.' Khoekhoe colexifies 'semen' with 'egg of frog' or 'egg of fish.' The Yoruba term *àtọ̀* contains *tọ̀* 'to discharge liquid,' while Burarra *burpur* also means 'mould,' and *gu-lol* gives rise to *-gulol* 'rotten.' Rotokas colexifies 'semen' with 'spit.' Ket *doˀŋ* is also the numeral 'three,' in Blackfoot, 'semen' is *okoyiim*, literally "his wolf," Lake Miwok *póṭa* also means 'foam' and 'be gray, be cloudy' (there is the optional complex term *ʔeláyni póṭa*, literally 'children foam,' for 'semen'), Nez Perce *cé·p* also means 'arrow,' 'bullet,' and 'cannon ball.' Wintu *kur* also means 'to be born,' 'to bear a child,' 'to be fertile,' and San Lucas Quiaviní Zapotec *mo'c* (< Span. *moco*) also 'phlegm.' Arabela *mashiquia* also denotes 'brushwood flowing downstream when the river is rising,' and Carib *apy tano* contains *apy* 'loin.' Guaraní *kuimba'e rykue* contains *kuimba'e* 'man,' and Lenakel *nɨpɨknisi-* appears to be analyzable as /nɨpɨk-nisii-/ 'tail-excrement-' (perhaps *nɨpɨk-* is also a colloquial designation for 'penis,' compare evidence in section 62 from other languages).

138. The Stomach

Representation: 97%
Motivated: 32.5%
Thereof Analyzable: 9.2% Thereof Colexifying: 23.6%
Thereof by Contiguity: 5.8% Thereof by Similarity: 4.7%
Recurrent associated meanings: guts/innards, womb, heart, eat/food, pregnancy, inside, crop, liver, feel/think, front, bag, middle, bladder, waist, faeces, chest, hypochondria, navel, diarrhea, liquid/water, big

Khoekhoe, Mbum, Kwoma, Ngaanyatjarra, Southeastern and Western Tasmanian, Badaga, Basque, Khalkha, Laz, Cheyenne, Chickasaw, San Mateo del Mar Huave, Kiliwa, Lake Miwok, Nez Perce, Wappo, Guaraní, Miskito, and Hawaiian colexify 'stomach' (or 'belly,

abdomen') with 'guts' and/or 'innards' more generally (note also Pipil *-tu:xih* 'stomach' in the Cuisnahuat dialect and *tuxih* 'intestines' in the Santo Domingo de Guzmán dialect, that Burarra *-buka* also denotes "part of intestines," as well as the presence of this association in Indo-European, mostly Slavic, Buck 1949: 254). The association is realized formally by noun class alternation in Swahili and by gender alternation in Embera. Ngambay has *kéy bò sìn* 'house big guts' (note also the redundant Kyaka term *anda-romba* 'house-belly/stomach'), and Bororo has *peguru kodobo-reu* 'guts type.of.basket-like,' alongside a term directly colexifying 'faeces,' as well as 'rest, residue' with 'stomach.' An association with 'faeces' is also found in Tsafiki (*hua pe-coló* 'big excrement-package'), and a semianalyzable term with 'faeces' is also found in the related language Cayapa. Piro has *hitš̲ka-mapa* 'excrement-bag/bladder' (compare the colexification of 'stomach' and 'bladder' in Nuuchahnulth and Hawaiian), while Hawaiian and Rotuman colexify 'stomach' with 'bag' among other meanings; this association is also found in some Indo-European languages, in particular Celtic and Germanic (Buck 1949: 253). Katcha, Yoruba, Kwoma, Toaripi, Badaga, Khalkha, Welsh, Carrier, Lesser Antillean Creole French, Nez Perce, Yaqui, Bislama, Hawaiian, and Rotuman colexify 'belly' and/or 'stomach' with 'womb' (also very common throughout Indo-European, Buck 1949: 252), and similarly, Buli, Rendille, Ngaanyatjarra, and Bislama with 'pregnancy' and/or 'be pregnant.' 'Stomach' and 'heart' are colexified in Ngambay, Abzakh Adyghe, Badaga (where the gloss for 'heart' is in quotation marks, suggesting that this usage may be figurative), Laz, San Mateo del Mar Huave (by the analyzable term *omeaats-aran* 'inside-INAL.POSS'), Itzaj (by the analyzable term *pusik'al* /puus-ik'-al/ 'dusting-wind-COLL'), Santiago Mexquititlan Otomí, Yuki, Arabela, and Guaraní. Analogously, Kiowa has *t'ę̨n-t'ǫ̨* 'heart-water' ('liquid' and 'stomach' are colexified in Bororo among other meanings, and there is a semianalyzable term featuring a constituent meaning 'water' in Kashaya). A complex term for 'stomach' on the basis of 'heart' is also found in Breton (Buck 1949: 253). Ngambay, Khalkha, and Yanomámi colexify 'stomach' with 'liver,' and Malagasy has *ambavafò*, containing *vàva* 'mouth' and *fò* 'heart;' note that there is a diachronic connection between 'stomach' and 'mouth' in Greek. Buin and Lesser Antillean Creole French colexify 'chest' (Buin 'chest of man' more specifically), and Basque 'bosom, breast' alongside 'gulf, bay.' Basque also colexifies 'belly, paunch' with 'navel,' and in Kyaka, there is a term that varies dialectally between these meanings. In Khoekhoe, there is a general term referring to internal organs of the trunk, including the 'stomach.'

Abzakh Adyghe and Itzaj colexifies 'stomach' with 'middle' (Baruya less generally with 'middle of the body,' Arabela similarly with 'center,' and Cashinahua with 'thick part in the center of something'), and Cavineña has *e-care-nani* 'INAL.POSS-middle/waist-hole.' The abstract notion 'inside' is colexified in Hausa, Dongolese Nubian, and Lake Miwok (note also the derived term in San Mateo del Mar Huave just mentioned above). Mirroring the Cavineña association with 'waist,' Cahuilla and Samoan 'stomach' with 'waist' directly.

Wintu colexifies 'stomach' with 'wrinkles' (and perhaps 'honeycomb tripe'), and there is a semianalyzable term where the identifiable constituent means 'wrinkled' in Wayampi. Lake Miwok, Arabela (by a semianalyzable term containing the classfier *-co* for receptables), Wayampi, Yanomámi, and Hawaiian colexify 'crop,' Ngaanyatjarra, Yir

Yoront, and Samoan ‘front (of person),’ and Basque and Bororo ‘hypochondria.’ Associations with ‘food’ or ‘eating’ are found in Oneida (*-khwalákhwa?*, containing the roots *-khw-* ‘food’ and *-l-* ‘be in or on’), Kaingang (*jẽn né* ‘eat container’), Tehuelche (*xa:t'en ~ xat'en*, derived from *xa:t'e* ‘to eat’ and colexifying ‘food.’ On the basis of this term, there is the redundant term *xa:t'en p'ate* ‘stomach/food skin;’ a semianalyzable term with ‘skin’ as the identifiable constituent is found in Berik), and Hawaiian (*pu‘u ‘ai-‘ai* ‘protuberance eat/food-RED’). Furthermore, Yir Yoront has the respect vocabulary term *maymay* reduplicated from *may* ‘food.’ Finally, Aguaruna and Rotokas colexify ‘stomach’ with ‘diarrhea’ (Aguaruna also with ‘cholera’), and in Kwoma, Ngaanyatjarra and Badaga, the ‘stomach’ is also considered the seat of emotions.

Other associations include: Buli *puuk* also denotes a particular clay vessel, and, in addition, colexifies ‘flower, blossom’ and ‘foam, lather.’ Hausa *ciki* also denotes the strips of cloth in a garment, and “a children’s aquatic game.” Khoekhoe features a term with very broad reference to ‘innards’ and ‘offals,’ including alongside ‘heart’ and ‘liver’ also ‘lungs’ and ‘kidneys;’ it also means ‘interior’ generally. Rendille *úur* also denotes the ‘character’ of a person. Baruya *munya* also means ‘above, high, top’ as well as ‘stone adze.’ Buin *moo-nogu* is analyzable as ‘dirt-like.’ The Dadibi term *moni hamago* contains *moni* ‘large,’ and Kyaka *romba* also denotes the “outer skin and fat layers over stomach.” Muna *taghi* is also used to refer to ‘lees’ or ‘sediment,’ and *randa* can also mean ‘bruised’ (also of fruits). Ngaanyatjarra *tjuni* also is used to refer to the “hollow of anything concave” and the “round part of a fruit,” while Nunggubuyu *mu̲lgu* also denotes the ‘stomach lining.’ Rotokas *kovapato* appears to be derived from *kova* ‘growth,’ ‘to grow, mature.’ Toaripi *luka* is also the name of a tree with hard timber, and Yir Yoront *pirrm* may also mean ‘body cavity’ as well as ‘the interior of house.’ Abzakh Adyghe *gᵒə* also means ‘surface, territory,’ Badaga *bevaru* is also a verb meaning “to spread grain out, to sort out grains,” and *karu* also means ‘dysentery’ and denotes “something that came from the stomach” inter alia. Basque *urdail* also means ‘rennet,’ Ket *hūj* also ‘room,’ and Khalkha *gedesy(n)* is also the name for “the loop for o/u in Khalkha script” and may be derived from *gede* ‘back of neck, occiput.’ Haida *k'iiji* also denotes the ‘swim bladder of fish.’ Welsh colexifies ‘calf,’ and Lesser Antillean Creole French *vant* also means ‘sale, auction.’ Tuscarora *útkweh* also means ‘groin’ and ‘pleura,’ and Central Yup’ik colexifies ‘gizzard.’ Arabela *sara-ca* appears to be analyzable as ‘uvula-CLASS.FRUIT.’ Chayahuita *anpopi-të'* is analyzable as ‘pith-CLASS.INSTRUMENT.’ The relevant Cubeo term shares its root with terms for ‘sweet potato’ and ‘alcoholic beverage made from sweet potato,’ and Guaraní colexifies ‘stomach’ with ‘spirit, conscience.’ Lenakel *tɨpweua* is also the name of a kind of breadfruit. Rotuman *taga* can also refer to a ‘pocket’ inter alia, while Samoan *manava* also means ‘smooth, soft side of a thing’ and, figuratively, ‘child.’ Sedang *pơtok* is also used to refer to the ‘abdomen’ of insects.

139. The Sweat

Representation: 75%
Motivated: 17.7%
Thereof Analyzable: 7.7% Thereof Colexifying: 10.1%
Thereof by Contiguity: 11.4% Thereof by Similarity: 3.2%
Recurrent associated meanings: heat/warmth, water/liquid/juice, steam

The most common of the relatively few lexico-semantic associations for 'sweat' (or 'perspiration') is that with 'heat/hot,' 'heated.' or 'warmth/warm' (as it is in Indo-European, Buck 1949: 263-264). These meanings are colexified in Buli, Hausa (alongside other apparently unrelated meanings), Badaga, Ket, Abipón, Aguaruna, Huambisa, Miskito, Fijian, and Samoan. Semianalyzable terms where a constituent with this meaning is present are found in Kolyma Yukaghir and Ineseño Chumash, Oneida has the verb *-ate?tukhwálha?* 'to sweat' revolving around the root *-?tukhwal-* 'to be hot' and Kiowa has *sʜdl-tʻǫ́ų* 'be.hot-water.' Unsurprisingly, Kiowa is not the only language to feature terms for 'sweat' related in some way to 'water,' 'liquid,' or even 'juice.' In fact, Toaripi has *maea ma ma* 'body water RED,' Abzakh Adyghe *pšʔentʔe ~ pšʔantʔe* contains elements meaning 'water' and 'violent,' Carrier has *nê-tsi-n-thû* '??-head-EPEN-water,' Santiago Mexquititlan Otomí *nxa̲nthe* /xa̲ni-dehe/ 'sprinkle-water,' which colexifies, like White Hmong, 'steam,' and Pawnee *kickaasisuʔ* is analyzable as /kic-kaʔas-his-uʔ/ 'liquid-break.out-PERF-NOM,' Lesser Antillean Creole French directly colexifies 'sweat' with 'water' (as well as 'rain'), and semianalyzable terms containing an element 'water' and/or 'liquid, juice' are found in Piro and Tsafiki.

Other associations include: Hausa *ji'bi* can also refer to a "large quantity," Kaluli *ha:fo:f* might be related to *ha:fó:* "deep breathing, panting, wheezing, asthma," Kyaka *pusi* also means 'cat' due to borrowing from Tok Pisin, and Ngaanyatjarra *parlulungu* means 'humidity, humid' and in the Northern dialect also 'sweat.' Kosarek Yale *wihin* 'sweat, tiredness' contains *wihi* 'ripe, big; tired.' Yir Yoront has *morr-ninn* 'body-sweat.smell.' Badaga *uri ~ huri* inter alia also means 'to flare up,' 'to fry,' 'jealousy,' and 'venom.' Basque *izerdi* also means 'dampness' and 'work,' while among the meanings colexified by Khalkha *kølysy(n) ~ kølesy(n)* is also 'payment for work, hire, fees.' Wintu *ƙiw* also denotes 'cooking stones' and a river rock which is put into acorn soup when cooking. San Mateo del Mar Huave colexifies 'sweat' with 'dermatitis,' while Central Yup'ik *uquryak*, a Hooper Bay and Chevak dialect term for 'heavy sweat' is analyzable as /uquq-yak/ 'oil-thing.similar.to.' Cashinahua *dabixtun* is also used to refer to 'grease on the body of a newborn.' Guaraní *ty'ái* appears to be analyzable as /ty-ai/ 'urine-line.' Sáliba *aixito* denotes also 'hot, boiling water' specifically, while Wayampi *pili?ay* is analyzable as /pili-ay/ 'smell-bad.' Bislama *swet* (< Engl. *sweat*) also means 'to put a lot of effort into something,' Kapingamarangi *hee* also 'where,' and Hawaiian *hou* also means 'new, fresh' and 'to push, thrust' inter alia. Samoan colexifies 'sweat' with 'wither' and 'waterfall.' Finally, Yay *haan*[5] also means 'promise.'

140. The Tear

Representation: 86%
Motivated: 41.6%
Thereof Analyzable: 37.8% Thereof Polysemous: 3.8%
Thereof by Contiguity: 34.2% Thereof by Similarity: 1.2%
Recurrent associated meanings: water/liquid/juice, eye, cry

Very frequently, terms for 'tear' are characterized by analyzability of the lexical type, with constituents meaning 'eye' (which sometimes colexifies further meanings such as 'face') and 'water,' 'liquid' and/or 'juice,' as in Itzaj *k'a' ich* 'juice/liquid eye.' Such terms are featured in Buli, Mbum, Berik, Abzakh Adyghe, Badaga, Ket, Sora, Kolyma Yukaghir, Carrier (where additional constituents are present: *nê-na-tsel-thû·* is analyzable as 'human-eyes-anus-water'), Upper Chehalis, Chickasaw, Haida, San Mateo del Mar Huave, Kashaya, Kiliwa, Lesser Antillean Creole French, Santiago Mexquititlan Otomí, Pawnee (in this language, the term is identical on the surface with that for 'jelly, jam,' but has a different underlying morphology), Pipil, Wintu, Yuki, Copainalá Zoque, Arabela (where the term is more precisely not of the lexical type, but derived by a classifier for liquids), Bororo, Embera, Guaraní, Hupda, Jarawara, Maxakalí, Miskito, Rama, Tsafiki, Wayampi, Bislama, Fijian, Hawaiian, Lenakel, Malagasy, White Hmong, Sedang, Tetun, and Yay. The pattern is furthermore etymologically detectable in Yoruba. There are, however, also other complex terms involving a constituent meaning 'eye:' Mbum has *ɓì-yâr* 'at-eye,' Toaripi *ovo-roro* 'eye-rubbish,' Yuki *hul-ḵat* 'eye-wet,' Rotuman *sui ne mafa* 'bone GEN eye' (the *sui* in this expression is thought to be a mere "doublet" of *sui* 'bone' by the lexicographer), and Bororo has an alternative term which, alongside constituents meaning 'eye' and 'water,' also features elements meaning 'thing' and 'fire.' Semianalyzable terms where the identifiable constituent is 'eye' are moreover present in Kanuri, Sentani, Yir Yoront, Highland Chontal, Kiowa, Xicotepec de Juárez Totonac, Tuscarora, Wintu, Kaingang, Sáliba, Hani, and Samoan.

Likewise, there are complex terms in some languages in which one constituent is 'water,' 'liquid,' or 'juice,' but the other one is not 'eye.' In this case, a verb meaning 'to cry' is the most frequent alternative. San Lucas Quiaviní Zapotec, for one, has *nnyi'ihs rùàa'n* 'water cries,' and such terms are also found in Efik and Yuki, while Bora has *máátyo-u* 'crying-CL.round' and Chayahuita *na'nëï'*, analyzable as /na'nërin-i'/ 'cry-CLASS.LIQUID.' Alternatively, Dongolese Nubian has *óñmɪssɛ* /óñ-míssɛ/ 'cry-sprinkle,' Koyraboro Senni, Ngambay, Buin, Upper Chehalis, and Aymara colexify the relevant meanings, and Kwoma has a semianalyzable term where the identifiable constituent is 'cry.'

Other complex terms involving 'water,' 'liquid,' or 'juice' include Khoekhoe has *ǁgam-ro-s* 'water-DIM-3SG.FEM' and Wappo *hu-méy* 'head-water.' Ancash Quechua and Manange colexify the relevant meanings, and semianalyzable terms where the identifiable constituent means 'water,' 'liquid,' or 'juice' are present in Kosarek Yale, Cayapa, Kaingang, Yanomámi, and Bwe Karen.

Other associations include: Hausa *k'walla* also means "repletion with fura" (a kind of dish), Koyraboro Senni *heeni* also means 'to cry' and "to make any loud or continous

noise" in general (an association also attested in Indo-European, Buck 1949: 1130-1131), Rendille colexifies 'tear' with 'seeds,' Kyaka with 'garden,' and Basque *negar* also means "lamentable, deplorable" and "dripping." *Malko* is also the name for a kind of small pear. Bezhta *maq'o* also means 'loom.' The Khalkha term *nilbusu(n) ~ nilmusu(n)* also may refer to 'mucus' or 'spittle,' Abipón *-aci-* also means 'ashes' and a root of the same shape also occurs in the term for 'tongue.' Carib *-enakulu* also means 'dirt in eye.' Cavineña *paanacaca* contains *caca* 'little,' Toba colexifies 'tears' with 'lung,' and Lesser Antillean Creole French *lam* also means 'spirit, soul,' presumably because of phonological collapse of Fr. *larme* and *l'âme*. Finally, Hawaiian *pūkai* 'lime bleach for hair, to bleach' rarely assumes the meaning 'salty tears;' this term may be related to *kai* 'sea.'

141. The Tendon

Representation: 75%
Motivated: 70.2%
Thereof Analyzable: 5.7% Thereof Colexifying: 64.5%
Thereof by Contiguity: 4.5% Thereof by Similarity: 0%
Recurrent associated meanings: vein/artery, nerve, muscle, thread/twine/fibre, root, gristle, string/cord, bowstring, line, flexible thing/elastic, sword, flesh

Very frequently, languages colexify 'tendon' with 'vein' and/or 'artery' cross-linguistically judging from the evidence of the sample. This is the case in Bakueri, Buli, Efik, Hausa, Dongolese Nubian, Yoruba, Baruya, Berik, Buin, Dadibi, Gurindji, Kyaka, Muna, Ngaanyatjarra, Nunggubuyu, One, Toaripi, Sahu, Kosarek Yale, Yir Yoront, Chukchi, Khalkha, Nivkh, Sora, Itzaj, Highland Chontal, Haida (for younger speakers only), Lakhota, Itzaj, Xicotepec de Júarez Totonac, Tuscarora, Arabela, Cashinahua, Cavineña, Cayapa (by a semianalyzable term containing an element meaning 'line;' 'line' and 'tendon' are colexified in Japanese and Biloxi), Chayahuita, Guaraní, Lengua, Miskito, Tehuelche, Toba, Yanomámi (by the analyzable term *mathõyãhi* /matha-yãhi/ 'leg-flesh,' compare Tetun *na'an-isin* 'meat/flesh-flesh' and note also that Kyaka and Toaripi have redundant complex terms with 'flesh' as the meaning of the additional constituent to single out the meaning 'tendon,' and that there is a semianalyzable term with the identifiable constituent meaning 'muscle, meat' in Hani), Bislama, Fijian, Great Andamanese, Hawaiian, Lenakel, Malagasy, White Hmong, Rotuman, Takia, and Yay. Moreover, Bora has a complex term for 'tendon' on the basis of 'vein' featuring, among other additional constituents, *méjpi* 'body,' Embera has *hir'ũkẽŋgúdroma*, containing *hir'ũ* 'foot' and *kẽŋgú* 'nerve, vein,' and Samoan has *uaua i soʻoga* 'vein/artery/pulse LOC joint.' Furthermore, there is the term *otsinuhyáhta? ahsliːyeˑ* in Oneida, containing *otsinuhyáhtu?* 'vein' and *ahsliːyeˑ* 'string, thread, yarn;' however, the term for vein is itself semianalyzable, containing the root *-nuhy-* for 'sinew.'

Many of the languages just mentioned, but also some others, colexify 'tendon' also with 'nerve.' This is the case in Buli, Hausa, Toaripi, Yir Yoront, Chukchi, Khalkha, Sora, Carrier, Ineseño Chumash, Itzaj, Lakhota, Nez Perce, Xicotepec de Juárez Totonac, Arabela, Aymara, Bororo, Cavineña, Chayahuita, Guaraní, Miskito, Ancash Quechua, Tehuelche, Yanomámi (again by the analyzable term *mathõyãhi* /matha-yãhi/ 'leg-flesh'), Hani, Ha-

waiian, and Rotuman (colexifying 'large nerve' specifically). Furthermore, Muna, Ngaanyatjarra, Biloxi, Upper Chehalis, Ineseño Chumash, Wappo, Guaraní, Jarawara, Lengua, Wichí, Yanomámi, Fijian, Hawaiian, Lenakel, Malagasy, and Tetun colexify 'tendon' with 'muscle' (Yanomámi and Tetun more specifically with 'muscular tissue'), Cheyenne, Haida, Itzaj, Lake Miwok, and Wappo with 'gristle,' and Aymara and Rotuman with 'flexible thing' and/or 'elastic' respectively (similarly, Ancash Quechua *anku* also means 'hard, difficult to break by stretching').

There is a further association, namely colexification with 'root,' which is found in Kwoma, One, Chickasaw, Ineseño Chumash, Pawnee, Jarawara, Lengua, Miskito, and Hawaiian (colexifying 'small root, rootlet' specfically). Gurindji colexifies 'single root of tree' specifically; furthermore, Kiliwa has a derived term, and there is a semianalyzable term where the identifiable constituent means 'root' in Highland Chontal. Khoekhoe, Tasmanian (Northeastern, Middle-Eastern, and Southeastern), Chukchi, Khalkha, Kolyma Yukaghir, Biloxi, Wintu, Yuki, Central Yup'ik, Lengua, and Bislama, by functional or provenience contiguity, colexify 'tendon' with 'thread,' 'twine,' and/or 'fibre,' Bislama also with 'fishing line' specifically. Finally, Upper Chehalis and Ineseño Chumash colexify 'bowstring,' Dongolese Nubian, Biloxi, Wintu, Copainalá Zoque, and Bislama 'string' more generally or 'cord,' and Japanese and Mandarin Chinese 'sword' (the association was borrowed along with the relevant term *jian4* from Chinese into Japanese).

Other associations include: Hausa *jijiya* also denotes a "tightening string of a drum," and Baruya *wɨrɨla* is also used to refer to "ridges on a shield" and "stringy fibres of wild yam." Muna *ue* is also denotes a unit of measurement inter alia. Ngaanyatjarra *pulyku*, in the Northern dialect, also denotes the "Pencil Yam." Abzakh Adyghe *le*, only perhaps meaning 'tendon,' otherwise also means 'be capable' and 'white, light.' Basque *zurda* means 'mane' and 'fishing-line' inter alia, and in the dialect of Lapurdi also 'tendon.' Khalkha *xujang* also means 'rheumatism, arthritis.' Haida *xay* also means 'warp,' Itzaj *jich'* also 'tighten' (another Itzaj term, *xich'el*, consists of a marker for inalienable possession as well as the root *xich'*, which can also mean 'wiry, thin'), and Nuuchanulth *łukłapt* is also the name of a spirit. Xicotepec de Juárez Totonac has a semianalyzable term containing an element meaning 'bone.' Tuscarora *unęhyá·θeh* also means 'plantain,' Wintu *la·* also 'skin' and 'stretch,' and Copainalá Zoque *nʌŋʌnʌŋʌ* also 'rope.' Aguaruna *wánus(e)* also means 'ankle bone,' Guaraní *tajygue* also 'strength' and 'lungs,' and Jarawara *kowisa-ri* appears to be analyzable as 'hurt-NMLZ,' while *habi* also colexifies 'tuber,' and, figuratively, 'courage' and 'audacity.' Imbabura Quechua colexifies 'tendon' with 'vine.' Bislama *string* also means 'erection, to have an erection,' and Fijian *ua* also 'wave, tide.' Figuratively, Hawaiian *aʻa* also means 'womb, offspring' (similarly, Lenakel *nouanul-*, containing *noua-* 'mouth, opening,' 'fruit' colexifies 'genealogical line, family, descent-group') as well as "to send greetings of love, joyous hospitality." Hani *saqguq* also denotes the 'pulse of a vein.' Finally, Sedang has *tróang húan* 'road grow/sprout/bud,' and Yay colexifies 'tendon' and 'vein' with "violin, fiddle."

142. The Testicle

Representation: 84%
Motivated: 47.4%
Thereof Analyzable: 21.4% Thereof Colexifying: 26.0%
Thereof by Contiguity: 10.3% Thereof by Similarity: 11.9%
Recurrent associated meanings: egg, scrotum, seed, round, fruit, ball, child/offspring, stone, penis, genitals, kidney, nut, energy

The 'testicles,' when expressed by motivated terms, are most frequently associated lexically with small roundish objects by metaphorical transfer. The findings on the basis of the present sample are largely in line with the results arrived at by Brown and Witkowski (1981) and Buck's (1949: 257) brief statement on Indo-European in particular, though some associations not present in Brown and Witkowski's study are uncovered here.

A frequent transfer pattern is that from 'seed' to 'testicle.' Fijian has *sore-ni-qala* 'seed-POSS-scrotum,' and derived terms, such as Chickasaw *intalop* /im-talop/ 'DAT-seed,' are also featured in Ineseño Chumash, Arabela, and Hawaiian, One has *amplu tala* 'male.genitals seed/round.thing,' while Welsh, San Mateo del Mar Huave, San Lucas Quiaviní Zapotec, Guaraní, Toba, Hawaiian, and Kapingamarangi colexify 'seed' and 'testicle' directly (Kapingamarangi 'seed of breadfruit' and Welsh 'pit of fruit' more precisely; Toba also colexifies 'bullet'), and there is a semianalyzable term in Kaingang. Like Kapingamarangi, an association with 'nut' and 'walnut' more particularly is present in Khalkha.

Moreover, White Hmong has *noob-qes* 'seed-egg,' and indeed, the association between 'testicle' and 'egg' is also common cross-linguistically. Sora, for one, has *'arre:-'kad-ən* 'egg-male.genitals-N.SFX.' An analyzable terms of the lexical type (with 'scrotum' acting as the contiguity anchor) is found in Dongolese Nubian, and derived terms are featured in Basque (where the 'testicle'-word is the plural of that for 'egg'), Blackfoot and Lengua, semianalyzable terms perhaps in Hupda, Macaguán, Lenakel, and Manange, and direct colexification occurs in Efik, Nganyatjarra, Carrier, Highland Chontal, Itzaj, Santiago Mexquititlan Otomí, Pawnee (with slight phonological deviations), the Cuisnahuat dialect of Pipil, Tuscarora, Bororo, Miskito, Piro, Ancash and Imbabura Quechua, Rama, Tsafiki, Yanomámi, Bwe Karen, and Lenakel; the association is diachronically recoverable in Wayampi. Moreover, Berik has *bol sui* /bola sui/ 'ball egg,' and similar associations on the basis of 'ball' are found in Japanese (*kin-tama* 'gold-ball'), Mandarin (*gao1-wan2* 'swamp/high/eminent/praise-globe/ ball/lump'), and Vietnamese (*hòn dái* 'ball genitals'), while 'ball' and 'testicles' are colexified in Nuuchahnulth and Bislama (compare the situation in English). Lenakel features a semianalyzable term where the identifiable constituent *noua* means 'fruit,' and associations with 'fruit' specifically are also found in Toaripi (*kō fare* 'scrotum fruit,' with *kō* also meaning 'breast, milk'), Samoan (*fuā-manava* 'fruit-belly'), and Tetun (*lasan-fuan* 'penis-fruit/heart'). 'Fruit' and 'testicle' are colexified in Yir Yoront (also with 'rolled string'), Cavineña, Bislama, and Hawaiian. A term bearing an association with 'stone' is featured in Ineseño Chumash (*is-xɨp* 'one's.own-stone'), a similar term is featured in Oneida, the association is realized by colexification in Welsh, and a

semianalyzable term where the identifiable constituent bears the meaning 'stone' is featured in Santiago Mexquititlan Otomí (note also the redundant Kanuri term *súsú tárwə́làm-bè* 'stone male.testicles-of'). 'Testicle' and 'kidney' are colexified in Khalkha and Rotuman. In fact, alongside the associations with specific smallish round objects discussed so far, there are also languages in which terms for 'testicles' make reference to 'roundness' directly, as already foreshadowed by the One term mentioned above. Rotokas has *orikoroisi* apparently containing *roi* 'intercourse' and the classifier *isi* for round objects, Basque has *barrabil*, containing *bil* 'round,' Nuuchahnulth *hupkimɬ*, which is presumably analyzable as /hup-qimɬ/ 'roundish.thing-CHUNK.SHAPED.OBJECT,' Wintu *siw*, which is is diachronically related to the word for 'eyeball,' Arabela, as alluded to above, has a term derived from 'sap, juice' by means of a classifier for round objects, Bislama *raon-wan* 'round-one,' Hani *davqwuv wuv-siq* 'scrotum RED-CLASS.ROUND,' and Vietnamese *tinh hoàn* 'energy round.object' (this association recurs in Bislama by colexification, where the relevant term is *paoa*, from English *power*). Hawaiian colexifies 'testicles' and 'small round object' directly, and Cayapa features a semianalyzable term with a constituent with precisely this meaning. As may have been noted from the discussion so far, 'scrotum' frequently acts as a contiguity anchor in complex terms of the lexical type. 'Scrotum' and 'testicle' are colexified furthermore in Buli, Koyraboro Senni, Swahili, Muna, Nunggubuyu, Aguaruna, Aymara, Cashinahua, Ancash Quechua, Bislama and Takia, and in Khoekhoe, the same root yields both 'testicle' and 'scrotum,' with different nominal designants disambiguating the referents. Another complex term involving a constituent meaning 'penis,' alongside the one in Tetun mentioned above, is Sko *òebi*, perhaps /òe-bí/ 'penis-floor/shell,' which colexifies 'testicle' with 'cheek.' Semianalyzable terms of this type are featured also in Kaluli, Sentani, Kosarek Yale, and Toba, while Tehuelche colexifies 'penis' and 'testicle' (for Sentani *mu haka*, where *mu* is 'penis,' compare *haka-bo-* 'run away'?). Finally, figuratively, 'testicles' may also be extended to "son under three years of age" in Great Andamanese and to 'offspring' in Hawaiian, while Guaraní has *ta'ỹi* /ta'y-i/ 'son/clot-DIM.'

Other associations include: Hausa *gwaiwa* also denotes "[t]he weight (clay or stone) at the end of the pole of an irrigation plant," Ngambay *gèm* also 'times,' and Rendille *jiláh* also 'burning coal.' Anggor *t̶imoef̶i* may also be capable of referring to a 'tree root,' and Kwoma *madii* also to "women's protruding labia." Badaga *oḍe* 'animal testicle' also means 'to break' and "wood, thicket, bush," while Greek *órchis* also means 'orchid.' Kolyma Yukaghir *önd'ed-abut* is analyzable as 'male-container,' while Upper Chehalis *má·čaɬ* also means 'pear.' Central Yup'ik *ingcu* is also used with the meanings 'nosebead' and 'mantle in gas lamp.' Bora *dómi-úúho* is analyzable as 'pubis-CL.chunk,' Tsafiki *pi'poca* might consist of *pi* 'water, liquid, juice' and *poca* 'cane of guadúa bamboo,' while Yanomámi *aruku* also denotes the 'eggs and larvae of wasps and bees.' Hawaiian *hua* also colexifies 'tuber,' 'produce, yield,' 'ovum,' as well as 'word, figure,' among other meanings, and the formally redundant reduplicate *huahua* also means 'fruitful, productive, prolific, have many children, lay many eggs' and is the name of a 'vulgar gesture.' *Kowaū,* another term in the same language, also means 'fish eggs.' Rotuman *ififi* can also be used with the meaning "in bunches or clusters." Bislama *bol* (< Engl. *ball*) may also refer to 'sac' and the

"soft belly of coconut crab," and *frut* also to "any individually sold item taken out of a packet."

143. The Tongue

Representation: 97%
Motivated: 20.6%
Thereof Analyzable: 3.1% Thereof Colexifying: 17.5%
Thereof by Contiguity: 11.2% Thereof by Similarity: 6.3%
Recurrent associated meanings: language/dialect/speech/word, blade, blade of oar, lick, trigger, foot of mollusk, mouth

'Tongue' is colexified with 'language,' 'dialect,' 'speech,' and/or 'word' in Hausa, Toaripi, Yir Yoront, Basque, Bezhta, Chukchi, Greek, Khalkha, Laz, Sora, Lesser Antillean Creole French, Quileute (here, the additional reading 'language' is rare), Cavineña, Embera, Rama, Bislama, and Hawaiian (this association is also common in Indo-European according to Buck 1949: 230, corroborating the commonness of the association in Eurasia detected in the sample, compare also Hilpert 2007), while Abzakh Adyghe has *bzeg°* /bze-g°ə/ 'language/area-surface,' and Tasmanian (Northeastern) has a semianalyzable term where the identifiable constituent means both 'mouth' and 'language' (which are often colexified as well, compare section 124).

Otherwise, Wichí *tok'ajlhech'e* contains elements meaning 'mouth' and 'egg,' and semianalyzable terms where the identifiable constituent is 'mouth' are attested in Nunggubuyu, Northeastern Tasmanian, Kiliwa, Wappo, and Hupda (note also that Kashaya *ha·ba* is etymologizable as **ʔaha-hiba·* 'mouth-tail'). There is just one language to colexify 'mouth' with 'tongue' (as well as 'manner of speaking') directly, namely Miskito. Hausa, Basque, Fijian, and Malagasy colexify 'tongue' with 'blade' (similarly, Chukchi, Fijian, and Hawaiian with 'blade of oar,' and Nivkh and Arabela with 'trigger'). Finally, Rama *kúup ~ múkup* contains *up* 'eye,' and Tsafiki *ni'caca* appears to be analyzable as /ni-ca'cá/ 'seed-eye.' Haida and Nuuchahnulth colexify 'tongue' with 'foot of (a certain) mollusk,' and Wintu and Manange colexify 'tongue' and 'to lick' (compare associations between 'tongue' and 'lick' in Indo-European, Buck 1949: 230).

Other associations include: Hausa *harshe* also means 'flame' and "[e]xtremity of a whip; point of a sword, knife, or loin-cloth." Ngambay *ndɔ̀n* also means 'to hunt,' Anggor *tefü* also "inside of pandanus fruit," and Burarra *ngarl* also denotes the 'taste of fresh meat or seafood.' Waris *minde* is also the name of a "pole with a hook for pulling fruit from a tree," while Basque *mihi* also means 'bit' and 'tap, flap' inter alia. Basque *mingain* also means 'masthead,' Bezhta *mic* also 'nettle,' and Greek *glṓssa* also denotes the 'sole' (*Solea solea*, compare German *Seezunge*). Ket *ējˀ* also means 'river island' and 'pine,' Khalkha *kele(n)* also 'bell clapper,' 'tongue of buckle' and "[a]nything resembling the tongue" generally, Japanese *shita* with different prosodic structure also 'down, under,' and Cheyenne *vétanove* also 'tongue of wagon' and 'tongue of shoe.' Haida *t'aangal* also means 'barb of fishhook,' and Itzaj *ak'* also 'vine.' Xicotepec de Juárez Totonac *sīˀmaˀkāˀt* also means 'uvula.' Central Yup'ik *alungun* is the name of a certain dog-feeding trough, and is derived from *alunga^{e}*

'home-made dog food' but also has the meaning 'tongue' in the Kuskokwim and Bristol Bay dialects. The root *-aci-* around which the Abipón term for 'tongue' revolves also means 'tear' and 'ashes,' in Cubeo, an identical root to that found in the term for 'tongue' yields the meaning 'bundle' when suffixed with another classifier, and Jarawara *abate* also means 'cheek.' Piro *n·u* may also be used with reference to a soft, fleshy object in general, while Wayampi *apɛkũ* also means 'reed of a clarinette,' and Yanomámi *aka* is also used to refer to the 'proboscis of the dipteros,' as well as to the 'starter of an outboard engine.' Bwe Karen colexifies 'tongue' with 'to poison,' Lenakel with 'fish,' and the Hawaiian term *alelo ~ elelo* inter alia also denotes "meat of the sea egg or sea urchin," *lelo* also "yellowish, especially the hue imparted to a whaletooth pendant ... by smoking," and Samoan *laulaufaiva* contains *lau*, meaning either 'leaf' or 'flat and thin object' generally, as well as *faiva*, which can mean 'fishing party,' 'job,' 'skill,' and 'business' inter alia.

144. The Tooth

Representation: 98%
Motivated: 22.5%
Thereof Analyzable: 3.1% Colexifying: 19.4%
Thereof by Contiguity: 4.1% Thereof by Similarity: 13.9%
Recurrent associated meanings: beak, cog, mouth, bite, claw of crab, tip/point, horn, thorn, ivory, jaw, sharp, blade, seed/grain of corn

Contiguity-based associations for this meaning (additional glosses such as 'tusk' or 'fang' are not reported in the below discussion) are relatively rare, and most recurrent patterns are metaphor-based extensions. Baruya, Gurindji, and Cahuilla colexify 'tooth' with 'mouth,' while Aymara has *laka ch'akha* 'mouth bone.' Moreover, Central Yup'ik has *keggun* /kegge-(u)n/ 'bite-device.for,' such a derived term is also found in Kolyma Yukaghir, while Abzakh Adyghe and Hawaiian colexifY nominal 'tooth' with verbal 'to bite' (Abzakh Adyghe also more specifically with 'to gnaw;' this association is also etymologically recoverable in Indo-European, Buck 1949: 231). Kyaka and Jarawara colexify 'tooth' with 'sharp' (and Maxakalí *xox* is also glossed as "sharp fragment"). Ngambay and Wappo colexify 'jaw' with 'molar tooth' specifically, and Hausa and Dongolese Nubian colexify 'tooth' with 'ivory.'

Turning to metaphorical extensions, Baruya, Kyaka, Itzaj, Arabela, Bora, Cashinahua, Lengua, and Rama colexify 'tooth' with 'beak' (Kyaka colexifies 'narrow beak' more specifically; note also Embera *kidhá* 'tooth' and *kidá* 'beak,' and compare section 5 for complex terms for 'beak' betraying this association), Basque, Greek, Khalkha, Welsh, Lesser Antillean Creole French, and Hawaiian with 'cog,' Kyaka, Chukchi, and Samoan with 'horn' (similarly, Takia *fai* colexifies 'horn' with "upper canine teeth" specifically, and also with 'crocodile'), Toaripi, Bislama, and Hawaiian with 'claw of crab,' Kyaka, Kosarek Yale, and Lengua with 'thorn,' Lake Miwok and Jarawara with 'blade,' and Itzaj and Wintu with 'seed' or 'grain of corn.'

Moreover, Toaripi, Kosarek Yale, and Jarawara extend 'tooth' to 'tip, point' in general (Toaripi also with 'summit' and 'corner'), and Baruya, Burarra, and Fijian colexify 'tooth' with 'edge.'

Other associations include: Efik *edet* also denotes a "gap in the edge of a tool" and "[t]he horns of a post between which a wall plate is laid." With the nominal designant *-s*, Khoekhoe (Haiǁom and Topnaar/ǂAonîn dialects) *apu-b* assumes the meaning "empty ostrich shell" which is used as a vessel. Hausa *hak'ori* also means "rough surface," as well as "[a]n embroidered edging of garments." Ngambay *ngange* may also refer to a 'limit, border.' Anggor *yahafɨ* seems to contain *yah* 'to say,' and Buin *kompe* also denotes "the bony toothridge in the beak of a hornbill or other bird." Kwoma *pu* also means 'conch shell, conch shell trumpet.' Kyaka *nenge* also means 'food,' and Muna *wangka* also means "put a wegde in wood" in verbal usage. Kosarek Yale colexifies 'tooth' with 'name,' Yir Yoront with 'upwards,' Abzakh Adyghe *ce* also means 'feather, fur,' 'tooth,' and 'defense,' and Basque *hortz* also means 'prong,' 'pick' and 'hook,' as well as 'share.' Greek colexifies 'tooth' with 'tine' and 'bit sticking out,' Japanese with 'leaf' (the meanings being distinct prosodically), and Khalkha with "the graph ... in old Khalkha script." Haida *ts'ing* also means 'pestle,' Lakhota *hí* also 'to arrive, reach, get here, to come,' and Lesser Antillean Creole French *dan* also 'prong.' The Pawnee root *aar* also means 'to become, to do,' and Bororo *o* also 'front.' Aguaruna *dái* is also the name of a bird species. Miskito *napa* also means 'harrow, rake,' 'needle of a sewing machine,' 'claw, fingernail,' and denotes the "tooth" of a zip. Piro colexifies 'barb' and 'fishhook,' and for Tehuelche *ʔor ~ ʔoːr ~ or ~ oːr*, note the similarity with *ʔor ~ ʔol ~ or* 'nose, beak, point of knife;' there is another similar sounding word meaning 'perhaps.' Toba colexifies 'tooth' with 'placenta.' Hani *seq* also means 'to lead, guide, take along' and 'to marry a woman.' Hawaiian colexifies 'tooth' with 'beak of octopus,' 'nipper of an insect,' 'Aristotle's lantern,' 'interlocking stones,' and patterns on a mat or tapa. A Samoan term for 'tooth' can figuratively also refer to an 'enemy,' and another term colexifies 'goods, supplies,' 'fortune, wealth,' and 'business.' The meaning 'tooth' for this term is restricted to polite usage. Manange *1sʌ* also means 'land, soil, ground, mortar,' while Sedang *hơnéng* appears to be derived from *héng* 'to have a toothache' by means of the nominalizing infix *<ơn>*. Bislama colexifies 'tooth' with 'sucker, shoot' of a plant.

145. The Urine

Representation: 74%
Motivated: 12.0%
Thereof Analyzable: 5.7% Thereof Colexifying: 7.7%
Thereof by Contiguity: 0% Thereof by Similarity: 2.7%
Most Commonly Associated Meaning: bladder, faeces, water, semen, gall, dung

Ngaanyatjarra, Tuscarora, Aguaruna, Huambisa, Yanomámi, and Bislama colexify 'urine' and 'bladder' (Anggor colexifies 'bladder' with 'to urinate' rather than 'urine'), while Ngambay has *kán-sìn* 'bladder/gall.bladder-excrement.' Dongolese Nubian and Copainalá Zoque colexify 'urine' with 'faeces' as well as 'dung,' which latter association is also found in Indo-European, particularly Celtic (Buck 1949: 274). The association is realized formally

by alternation of nominal designants in Khoekhoe. Similarly, the Kashaya term contains an element meaning 'to excrete,' and there is a semianalyzable term for 'urine' featuring an element meaning 'faeces' in Sko.

Ngaanyatjarra and Fijian colexify 'urine' with 'gall' (Ngaanyatjarra also with 'gall bladder'). Hawaiian has *hana-wai* 'work-water,' which also colexifies 'irrigation' and 'menstruation,' Welsh *dŵr (o'r bledren)* 'water (of bladder),' and Kiowa *sʜ'-tsoue* 'burst-water' (though the connection with the verb is somewhat unclear; the connection with 'water,' on the other hand, is also attested, though weakly, in Indo-European, evidenced by cognates in Ancient Greek and Sanskrit with the respective meanings, Buck 1949: 273). Semianalyzable terms with a constituent meaning 'water' or 'liquid' are found moreover in Kiliwa, Cayapa, Rama, and Tsafiki; there are also redundant terms featuring a verb meaning 'to urinate' as the second constituent. Tuscarora has the term *uʔnhęhsú·kreh*, underlyingly /u-ʔnhęhs-ukr-eh/ 'NOUN.PREFIX-egg/testicle-rubbish-NOUN.SUFFIX,' for "foul or disgusting urine; a slovenly or slatternly person so fithly as to emit an odor of urine," the stem of which also yields the term for 'semen.' Khoekhoe colexifies 'semen' with 'urine' by an archaic term, and Kwoma has *moku sobo* 'semen raw/unripe/pure' (*sobo* colexifies also other meanings).

Other associations include: Efik *i'kïm* can also refer to "[a]ny disease of the urinary organs," Hausa *k'ura*, meaning 'urine' and "drinking water in which nothing has been admixed" in the dialect of Sokoto also means 'dust,' and *fitsari* also to 'humiliate a person in public.' Yoruba has *ì-tọ̀* 'NMLZ-discharge.liquid,' Baruya *suya* also means 'tail,' and Buin *iu* also 'honey' (the word is also a toponym for a mountain and a river). Burarra *darrjala* (which is also the name of a particular tree) might be related to *darrja* "be hot, as the sun shines hot," and Kyaka *puu* also denotes the "Calamus or lawyer vine" and a 'thread, rope' among other meanings. Yir Yoront *kachl* also means 'gill of fish,' Haida colexifies 'stale urine' with 'urine vessel, peepot' and 'amniotic fluid,' and Pawnee *asuris* is also used with reference to the 'odor of urine.' Central Yup'ik *teq'uq ~ etquq* might be analyzable as /teq-quq/ 'anus/bottom-one.like.' Arabela *shaaca* also means 'insipid, flavorless,' Aymara has a term for 'fermented urine' which in fact also means 'fermented,' while Bororo *iku-ru* seems to be analyzable as 'line-fire.' Guaraní *ty* also means 'juice' and 'big pile of something,' and Jarawara *yoka* also means 'athlete's foot.' Ancash Quechua *ishpay* also means 'trench of large trees on small plants,' and *pichi* is also an indefinite pronoun. Toba *lte* also denotes 'impurities in water' and 'characteristic color, characteristic sign, mark.' Bwe Karen *ʃi* also means "to be small, little; younger" and 'to dazzle,' Fijian *mi* also "to run in a small stream," and Hawaiian *mī* also means 'to dream' inter alia. Finally, it should be noted that Haida distinguishes lexically between 'fresh urine' and 'stale urine,' and Wappo between urine from males and females.

146. The Uvula

Representation: 27%
Motivated: 60.8%
Thereof Analyzable: 45.8% Thereof Colexifying: 15.0%
Thereof by Contiguity: 21.3% Thereof by Similarity: 38.8%

Recurrent associated meanings: tongue, throat, child/son, vagina/clitoris, tonsil, little, neck

The 'uvula' is frequenly designated by complex terms of the lexical type constituting a metaphorical transfer from other body-parts or entities, with 'tongue' acting as the contiguity anchor. Among those, one pattern recurs cross-linguistically, namely that from 'child' or 'son' more specifically, as in Khalkha *keyken kele* /keyken kele(n)/ 'child tongue.' Such a term is also featured in Chickasaw, and the association is by colexification in Xicotepec de Juárez Totonac (the term also colexifies 'brood of animals,' 'loop of a net,' and 'interest'). Note also Tetun *nanarak-oan* 'palate-son.' Other complex terms are Khoekhoe *ǂkhari-nam-s* 'small-tongue-3SG.FEM,' Badaga *kiru na:lange* 'little tongue,' Sora *əkantalaŋən* /ə-'kanta:-'laˑŋ-ən/ 'POSS-branch-tongue-N.SFX,' and *le:r'laŋən* /lˀeːr-'laˑŋ-ən/ 'grow.in.size-tongue-N.SFX,' Welsh *tafod bach* 'tongue took/hinge' and *tafod-ig* 'tongue-DIM,' Central Yup'ik *alungutayaaq* /alungun-taq-ya(g)aq/ 'tongue-device.for-little,' Guaraní *ape-kû-guy* 'skin-tongue-behind' (there is also the alternative term *ape-kuatî* 'skin-squirrel'), Yanomámi *akathamɨkɨ*, consisting of *aka* 'tongue,' *thamɨ* 'internal part of body part' and the quantal classifier (see § 4.4.1) *kɨ* (this term coleixifes 'velum' and 'gill'), Fijian *yame-leka* 'tongue-short,' Hani *lalngavq*, presumably /lalma-ngavq/ 'tongue-get.stuck,' and Yay *lin*^{6} *kay*^{2} 'tongue thing.' Finally, Itzaj has *ak' kal* 'tongue neck/throat' alongside *t'uy u-kal* 'vagina 3SG.POSS-neck/throat,' and in fact, this is another major recurrent association. Terms involving constituents meaning 'throat' and 'vagina' or 'clitoris' more specifically are also found in Carrier and Kiowa, and the association is by colexification in Wintu. 'Throat' and 'uvula' are colexified in Muna and Tuscarora; there is a semianalyzable term involving 'throat' in San Mateo del Mar Huave. Nez Perce has *múˑx̣s-ṅes* /múˑx̣s-eʔs/ 'swallow-INSTR,' and there are semianalyzable terms betraying this association in Wintu and Piro. Finally, 'uvula' is colexified with 'tonsil' in Blackfoot, Badaga, and Lesser Antillean Creole French.

Other associations include: Hausa *beli* also denotes the "[a] small bud-like growth at the joints of corn-stalks," and is the name of various diseases inter alia. *Haki-n wuya*, another Hausa term, is analyzable as 'grass-GEN neck.' There is also another semianalyzable term where the identifiable constituent means 'neck' which also denotes a 'retropharyngeal abscess.' Basque *aho-gingil* is analyzable as 'mouth-lobe.' Greek colexifies 'uvula' with 'grapes,' while Haida colexifies 'uvula' with 'esophagus.' Embera features a semianalyzable term involving a constituent meaning 'way,' and Samoan *alelo* is also a term for the "eyes of a snake or eel."

147. The Vein

Representation: 80%
Motivated: 65.0%
Thereof Analyzable: 20.0% Thereof Colexifying: 46.8%
Thereof by Contiguity: 5.5% Thereof by Similarity: 17.9%
Recurrent associated meanings: tendon, nerve, blood, way/street, muscle, root, thread/string/fibre, pulse, gristle, line, lode, liana, rope, fishing line

The 'vein' is frequently colexified with 'tendon, sinew' (see also section 141). This is the case in Bakueri, Buli, Efik, Hausa, Dongolese Nubian, Yoruba, Baruya, Berik, Buin, Dadibi, Gurindji, Kyaka, Muna, Ngaanyatjarra, Nunggubuyu, One, Toaripi, Sahu, Kosarek Yale (marked with a question mark in the source), Yir Yoront, Chukchi, Khalkha, Nivkh, Sora, Highland Chontal, Haida (for younger speakers only), Lakhota, Itzaj, Xicotepec de Juárez Totonac, Tuscarora (colexifying 'plantain'), Arabela, Cashinahua, Cavineña, Cayapa (by a semianalyzable term containing a constituent meaning 'line'), Chayahuita, Guaraní, Lengua, Miskito, Tehuelche, Toba, Yanomámi (by the analyzable term *mathõyãhi* /matha-yãhi/ 'leg-flesh'), Bislama, Fijian, Great Andamanese, Hawaiian, Lenakel, Malagasy, White Hmong, Rotuman, Takia, and Yay. Furthermore, Abzakh Adyghe has *ʎəntfe*, containing *ʎə* 'blood' and *fe* which perhaps bears the meaning 'tendon,' Miskito *tala wayka* 'blood ligament,' and there are semianalyzable terms featuring an element meaning 'blood' in Haida, Oneida, and Hani. Due to the association with 'tendon,' 'vein' is also colexified with 'gristle' in Welsh, Highland Chontal, and Haida, and with 'line' in Buli, Dongolese Nubian, and Bislama.

In general, associations are quite similar to those for 'tendon, sinew.' Buli, Hausa, Swahili, Toaripi, Yir Yoront, Basque, Chukchi, Khalkha, Sora, Itzaj, Lakhota, Xicotepec de Juárez Totonac, Arabela, Cavineña, Chayahuita, Embera (associated with different genders), Guaraní, Kaingang, Miskito, Tehuelche, Yanomámi (again, by the analyzable term *mathõyãhi* /matha-yãhi/ 'leg-flesh'), Hawaiian, and Rotuman colexify 'vein' with 'nerve.' Muna, Ngaanyatjarra, Waris, Welsh, Guaraní, Lengua, Fijian, Hawaiian, Lenakel, and Tetun colexify 'vein' with 'muscle' (and Yanomámi with 'muscular tissue'), and there might be a semianalyzable term on the basis of 'muscle' in Rama.

Rendille, One, Basque, Carib, Lengua, and Miskito colexify 'vein' with 'root' (similarly, Gurindji more specifically colexifies 'single root of tree,' and Hawaiian 'small root, rootlet'); moreover, Kanuri has *zâr bû-bè* 'root blood-of,' and Bora *bájkyemóóho* /bájkyeémóóhou/ 'root-liana.' Analogously, Copainalá Zoque colexifies 'vein' with 'liana,' and Tsafiki has *a'sán silí* 'blood liana/string.'

Chukchi, Kildin Saami, Lengua, and Bislama colexify 'vein' and 'thread, string' or 'fibre,' while Yaqui has *ojbo wii'i* 'blood thread.'

In fact, one difference between terms for 'tendon' and 'vein' is, as has emerged from the previous discussion, the frequent presence of 'blood' as a second constituent in complex terms for the latter acting as a contiguity anchor, as seen in Abzakh Adyghe and Miskito for the association with 'tendon,' in Kanuri for the association with 'root,' and in Yaqui for the association with 'thread.' A major difference with respect to 'tendon,' in spite of the many similarities, is also the presence of a transfer from 'way, street' to 'vein,' also with 'blood' acting as a contiguity anchor, as in Huambisa *numpa jinti* 'blood way.' This pattern is common in South America among the sampled languages, occurring alongside Huambisa in Aguaruna, Cavineña, Toba, and Wichí, but also attested in Mbum, Kolyma Yukaghir, Chickasaw, Santiago Mexquititlan Otomí, Malagasy, and Sedang. There are also some other complex metaphor-driven terms where one constituent is 'blood:' Hausa has *igiyar jinni* 'rope blood,' Ket analogously *sulaŋ* /sūl-àŋ/ 'blood rope' ('vein' and 'rope' are colexified in Copainalá Zoque), Japanese has *kekkan*, perhaps analyzable as /ketsu-kan/

'blood-conduit,' Biloxi *haiti´* /hai-ti/ 'blood-house,' Kiliwa *khwat=h+yuul* 'blood=3+flow,' and Mandarin *xue4-guan3* 'blood-pipe.' Semianalyzable terms are found in Pawnee (which colexifies 'rubber') and San Lucas Quiaviní Zapotec.

Otherwise, Greek, Khalkha, and Embera colexify 'vein' with 'lode' (Greek figuratively also with 'talent'), and Lavukaleve and Bislama with 'fishing line.' Khalkha, Nez Perce, and Samoan colexify 'vein' with 'pulse,' and Vietnamese has *tĩnh mạch* 'calm pulse.'

Other associations include: Buli *jiin* is also the name of a "tightening string of a drum,' Hausa *majinaciya* appears to contain *majina* 'mucus, snot' and also denotes a "weed with red fluid in it," Rendille *híy* also means 'sour milk' and 'relatives, kin,' while Baruya *wɨrɨla* also denotes "ridges on a shield" and "stingy fibres of wild yam." Dadibi *kigibili* is presumably analyzable as /kigi-bilibo/ 'maggot-walk.' Muna *ue* is also a unit of measurement inter alia, and Ngaanyatjarra *pulyku* (Northern dialect) also denotes the "Pencil Yam." Badaga *kuḍi* also means 'sprout, shoot' and 'penis' inter alia, Basque *zain* also "core, gist, crux" inter alia, Khalkha *sudal* also 'rings of a tree,' and 'ridge, stripe' inter alia, *xujang* also "rheumatitis, arthritis," Welsh *gwythïen* also 'seam,' Ineseño Chumash also 'bowstring,' Kiowa *kʻiн* also 'porcupine,' 'fire,' and 'to be heavy,' Lake Miwok *ċííkem* also 'to be striped,' and Itzaj *wich'el* contains *wich'*, which can also mean 'wiry, thin.' Lesser Antillean Creole French *venn* also means 'seam.' Cubeo *pɨ̃pɨme* also means 'spiderweb' and the root *pɨ̃pɨ* indeed yields the meaning 'spider' when suffixed with a different classifier. Guaraní *tajygue* also means 'strength' and 'lungs,' and Piro *kotsa* also 'leaf' and 'child who is thin and underdeveloped.' Bwe Karen colexifies 'snake,' Hawaiian *aʻa* may figuratively also refer to 'womb, offspring' as well as "to send greetings of love, joyous hospitality; joy at greeting a loved one," and Lenakel *nouanul-*, containing *noua-* 'mouth, opening, fruit,' may also refer to a "genealogical line" and a 'family, descent-group.' Tetun *uat* also is used with the meaning 'rain of wood,' Bislama *string* also means 'erection, to have an erection,' and Fijian *ua* is also used with the meaning 'wave, tide.' Finally, Yay colexifies 'tendon' and 'vein' with "violin, fiddle."

148. The Womb

Representation: 64%
Motivated: 67.6%
Thereof Analyzable: 48.6% Thereof Colexifying: 23.3%
Thereof by Contiguity: 22.8% Thereof by Similarity: 29.8%
Recurrent associated meanings: child/son/daughter/baby, stomach, house, placenta/afterbirth, pregnancy, bag, place, netbag, guts, embryo/foetus, nest, receptable, sit, heart, vagina/vulva, bladder, basket, palace, mother, give birth

Terms for 'womb' (or 'uterus') are frequently metaphor-driven, with 'child,' or more specifically 'son' or 'daughter,' acting as the contiguity anchor. Terms with 'house' as the source concept, as in Yaqui *asoa-kari* 'daughter-house,' are found in Mbum, Ngambay, Yoruba (where 'embryo' rather than 'child' is the contiguity anchor), San Mateo del Mar Huave (where 'house' is colexified with 'nest,' note also Berik *tane gol* 'child nest,' that Bezhta colexifies 'womb' with 'nest' alongside 'bear,' and that Cashinahua colexifies

'womb' with 'hole'), Cavineña, and White Hmong (see also Matisoff 2008: 176 for Tibeto-Burman specifically). Similarly, Japanese has *shi-kyū* 'child-palace' and Mandarin *zi3-gong1* 'child palace,' the Japanese term being calqued. Alternatively, Central Yup'ik has *enliaq ~ neliaq* /ena^{e}-liaq/ 'house-made,' and Kyaka *romba anda* 'stomach house/nest' alongside *ingi anda* 'intestines house/nest' and *wane yakera petenge anda* 'child embryo sitting house/nest' (note that Kyaka *anda* also colexifies 'nest,' and hence this association is also present here). The association with 'stomach' and 'child' as contiguity anchor recurs in Baruya (*bwaraminya* /bwaranya-munya/ 'baby-stomach') and Vietnamese (*dạ con* 'stomach child'); very similarly, Mbum has *gûn-bîl* 'child-abdomen' (see also Hilpert 2007 for this association). Furthermore, Dongolese Nubian has *kumátté-n-tu* 'vulva-GEN-stomach/interior' ('womb' and 'vagina' are colexified in Carib), Meyah *ojóna otkonú* 'married.woman stomach,' and Ket *āmd huūj*, which contains elements meaning 'mother' and 'stomach' (compare also complex terms in Indo-European with a constituent meaning 'mother' reported by Buck 1949: 255, as well as Bislama *basket blong mama* 'basket POSS mother/pregnant,' and, in turn with the Bislama term, compare Fijian *kato-ni-gone* 'basket/box-POSS-child' and the colexification of 'womb' and 'basket' in Lenakel). 'Womb' is colexified directly with 'stomach' and/or 'belly' in Katcha, Yoruba, Kwoma (colexifying also 'chest'), Toaripi, Badaga, Khalkha, Welsh, Carrier, Lesser Antillean Creole French, Nez Perce, Yaqui, Bislama (also with 'pregnancy' and 'be pregnant' in Bislama and one of the Khalkha terms), Hawaiian, and Rotuman (see also Buck 1949: 255 for evidence connecting 'womb' and 'stomach' in Indo-European). Moreover, Efik colexifies 'belly of animal' specifically (by a term that may be derived from a verb meaning 'to conceal'), and in Miskito the association is present, but rare.

Returning to complex terms with 'child,' the Kyaka association with the notion of 'sitting' is also present in Kiliwa (*mnyis=waa-u?* 'foetus=sit-OBL,' note that Mandarin and Khalkha colexify 'foetus' and 'womb' directly and compare the Yoruba term with a constituent meaning 'embryo' mentioned above). Khoekhoe has *ǂnû-ǃgau-s* 'sit-be.left.over-3SG.FEM.' In Badaga, a term for 'womb' is *ku:su pae* 'baby bag.' Similar terms on the basis of 'bag' as the source concept are also found in Kiowa and Piro, while Swahili has *fuko la uzazi* 'bag of birth' (alongside *chupa ya uzazi* 'bottle of birth;' a term based on a verb meaning 'to give birth' is also featured in Hausa and Toba). Hawaiian colexifies 'womb' with 'bag,' and there is a semianalyzable term in Highland Chontal (relatedly, Kildin Saami *pūŋŋ* is probably a loanword from Norwegian *pung* 'bag'). An alternative Badaga term is *gabba pae* 'pregnancy/happy.event bag,' and 'womb' is colexified with 'pregnant, pregnancy' in Baruya, Khalkha, Bislama, and Rotuman. Piro has *whenewlu mapa*, containing *whene* 'child' and *mapa*, which colexifies 'bag' and 'bladder;' 'bladder' specifically is colexified with 'womb' in Hawaiian. Sora has *dərakkʊ:'onən* /dərakkʊ:-'o:n-ən/ 'vessel/receptable-child-N.SFX.' Such terms are also found in Kashaya and Hawaiian, and the association is mirrored only somewhat differently in Toba (*l-co'o-oxo-qui* '3SG.POSS-give.birth-NMLZRECEPTABLE/ENCLOSED.SPACE). A semianalyzable term where the identifiable constituent is 'receptable' is also found in Yanomámi. Moreover, Efik has *ëbiët ëyën* 'place child,' and such terms are also found in Kanuri, Basque, Arabela, and Tetun.

There is also a wealth of complex terms with one constituent meaning 'child,' but the semantics of the other constituent being not among those discussed so far. Kolyma Yukaghir has *uon-könme* 'child-friend,' Chickasaw *oshaatoba'*, analyzable as /oshi' aa-toba-'/ 'son DAT-become-NMLZ' and *oshạ'to'*, analyzable as /oshi' a̲lhto-'/ 'son be.inside-NMLZ' (this term is archaic), Haida *gid dllt'iisra*, containing *gid* 'child' and *t'iis* 'be in contact,' Lake Miwok *ʔélay ṣúkúhni*, containing *ʔélaj* 'child' and *ṣúkuh* 'stay, remain,' Oneida *-wilalahkwaʔ*, containing the roots *-wil-* 'baby' and *-l-* 'be in or on,' Santiago Mexquititlan Otomí *nt'oxu̲bätsi*, which is analyzable as /nt'ots'i-bätsi/ 'granary-child,' Guaraní *memby-ruru* 'son-swollen,' Hupda *tæ̃ʼh-yud* 'child-clothes' (marked with a question mark in the source), Kapingamarangi *mee dugu dama* 'thing put baby,' and Manange *1kola1ʈupʌ4kʰja*, containing *1kola* 'child,' *1ʈu* 'stay,' and *4kʰja* 'place.' Furthermore, Laz, Upper Chehalis, and Lake Miwok have derived terms from 'child,' and Badaga and Welsh colexify 'womb' with 'heart,' with the additional meaning being obsolete in Welsh.

Associations exclusively realized by colexification in the languages of the sample are: Hausa, Khoekhoe, Ngaanyatjarra, Sko, Khalkha, and Cahuilla colexify 'womb' with 'placenta' or 'afterbirth' (Hausa also with 'parents' and 'birthplace,' and Khalkha also with 'place to lie down, cave, den, lair'), Sahu, Badaga, and Sedang with 'guts,' and Buin, Burarra, Kwoma, and Takia with 'netbag' (Kwoma also with 'hill' and 'mountain,' inter alia).

Other associations include: Hausa *mahaifa* also means 'parents,' while Khoekhoe *ǁhās* also means 'ravine, gorge' and 'gully,' and Muna *tie* also "litter, time of giving birth (of animals)." Badaga *karu* also means 'heart' (in quotation marks in the source), 'dysentery,' "something that came from the stomach" and other things, and Khalkha *xabisu(n)*, apparently derived from *xabi* 'vicinity, neighborhood,' also means 'rib' and "trimming or metal plates on the bottom of a coat or mail." Nez Perce *ʔilúˑt* may also refer to the "side, the part over the ribs of animals," and the Nuuchahnulth term *tiičsy̓aapi* is analyzable as /tǐˑč-sy̓i-api/ 'alive-thing-stand' and colexifies 'life,' 'life principle' and 'childbirth.' Pipil (Cuisnahuat dialect) *xina:ch* also means 'ovary' and 'egg in chicken.' Tuscarora *yętʔnęhtʔáhstha?* contains *-'nęhT-* 'to bury' (this term is marked as being unclear in the consulted source), Lengua *tathnak* might contain *tathna* 'navel,' Miskito *plauya* may rarely also refer to the 'bladder,' and Great Andamanese *ôtârain* seems to be derived from *ârain* 'gurjon tree' (*Dipterocarpu*s sp.). Hawaiian *pūʻao* also means 'mesh of mats,' *ʻōpū* also 'tendon, vein, muscle' (the meaning 'womb' is said to be figurative in the source), 'crop of bird,' and 'disposition,' and *pupuʻu* "to double up, draw the limbs together," hence 'foetal position,' and hence also 'womb' inter alia (this term may be derived from *puʻu* with the basic meaning 'protuberance'). Sedang *klea* also denotes the "inner edge of bamboo or of kɷmea square strip of rattan." Samoan *faʻa-ʻautagata* is analyzable as 'CAUS-be.shorthanded,' and Sedang *xoa* also means 'chest.' Tetun *knotak* also can refer to the 'waist.'

149. The Wrinkle

Representation: 48%
Motivated: 22.9%
Thereof Analyzable: 8.1% Thereof Polysemous: 14.8%
Thereof by Contiguity: 4.6% Thereof by Similarity: 15.5%
Recurrent associated meanings: crease/fold/pleat, ripple, skin, crumple, wither

Recurrent associations for this meaning are very few. There are some complex terms of the lexical type, where one of the constituents is 'skin,' with varying semantics of the second element. Abzakh Adyghe has *ɜ́ᵊe-λe-r* 'skin/surface-be.located-SUFFIX' which also denotes a 'line' generally, and Piro *tslaha-mta* 'screen/grate/bars-skin' for "fine wrinkles in skin (not from age)." Furthermore, Koyraboro Senni has *kuurukuuru*, reduplicated from *kuuru* 'skin.'

Khoekhoe, Greek, Khalkha, Itzaj, Central Yup'ik, and Lesser Antillean Creole French colexify 'wrinkle' with 'crease,' 'fold,' and/or 'pleat,' Khoekhoe and Rotuman with 'crumple,' as of clothes, Greek, Welsh, Nunggubuyu, Bororo, and Fijian with '(have) ripple(s)' (similarly, Hausa colexifies 'wrinkling in water after long continuance' as well as 'emaciation,' while in Fijian the meaning colexified is 'wrinkle on brow' more precisely). Kosarek Yale and Yay colexify 'wrinkle(d)' with 'withered' (Kosarek Yale also with 'to become loose,' said of the hide of a drum).

Other associations include: Efik *ufrä* also denotes a 'leaping, jump,' and figuratively, "wrinkling, corrugating" among other meanings. The root of the Khoekhoe term for 'wrinkle,' *ǁai ~ ǁairo ~ ǀhō*, also means 'to shrivel of wet paper' as a verb. Yoruba *ì-wunjọ ~ ì-hunjọ* is analyzable as 'NMLZ-to.wrinkle.or.shrink,' and Gurindji *wanyjarrng* contains *wanyja* 'wrinkled yam.' Kwoma colexifies 'root,' while Kyaka *moo* can also refer to a mark or indentation in sand or metal. Nunggubuyu *–narmanarma-* also means 'to be furrowed.' Tasmanian terms in all varieties except the Northern for which data are lacking are said to also mean 'wart,' 'scar,' and 'tail,' and Basque *zimur* is also used with the meanings "empty chestnut," "ungrateful, thankless," and "tight, mean, stingy." Ket *kuraŋbet* contains *kud* 'bend' and *bed* 'make.' The Itzaj term *otz'tik* may also be used to refer to 'foam' or 'froth,' Lesser Antillean Creole French *pli* also means 'more, most,' Nez Perce *yuḱúˑmyuḱuˑm* also 'crumpled,' while the Tuscarora root *-θriʔr-* yields both terms meaning 'wrinkle' and 'snail,' depending on the noun suffix attached. Wintu colexifies 'wrinkled' with 'stomach' and perhaps 'honeycomb tripe.' Central Yup'ik *imegglug-* is analyzable as /imeg-rrluk-/ 'roll.up/fold.up-one.that.has.departed.from.its.natural.state-,' *qelengllak* also means 'scar, kink,' and *qacu-* also 'to be loose' and 'to sag.' Guaraní *chaʼĩ* also means 'bad' and 'ugly,' Miskito colexifies 'to wrinkle' with 'to rumple,' Wayampi *kala* also means 'rough' inter alia, Bwe Karen *θitru* also 'crumpled,' Lenakel *uɬikuɬik* is reduplicated from *uɬik* 'tough' (as of meat and other food), and Hawaiian *minomino* is reduplicated from the base *mino* 'dimple, depression, dent' and also means 'messed' with reference to dresses.

150. The Dawn

Representation: 86%
Motivated: 51.0%
Thereof Analyzable: 33.8% Thereof Colexifying: 17.1%
Thereof by Contiguity: 30.7% Thereof by Similarity: 8.6%
Recurrent associated meanings: light/light up/bright, day, morning, sunrise, night, come/arrive, sun, twilight, clear, dusk, tomorrow, cock crow, dark, white/become white, morning prayer, greeting, eye, land, mouth

The most frequent association is, unsurprisingly, that with 'day,' occurring in a variety of subtypes. 'Day' or '(be) day' and 'dawn' are colexified directly in Baruya, Wayampi, and Hawaiian. As for complex terms, 'light,' 'light up,' or 'bright' is the most common meaning of the second constituent (a meaning also commonly associated with 'dawn' in Indo-European according to Buck 1949: 993), as in Ket *kūn* /kəˀn-iˀ/ 'bright-day' or Carib *emamɨlɨ*, derived from *emamɨ* 'be light.' Analyzable terms of the lexical type as in Ket are also found in Buli, Bislama and Tetun, and derived terms similar to that in Carib also in Abipón, Bora, and Jarawara. There are also variants of this pattern: Ngambay has *tà lò àrɛ* 'mouth time/day bright/lit.up' (with *lò àrɛ* being a term for 'day' itself), Kyaka *yuu nombalo* 'day first.light.of.day,' dialectal Basque *argi-haste* 'light-beginning,' Khalkha *sira gere* 'yellow light' and *gegegere-*, which is an inchoative verb derived from *gegere(n) ~ gege(n)* with the basic meaning 'daylight, morning daylight,' Itzaj *saska'tal* /sas-kab'-tal/ 'bright-world-come' (and variants of this term, including *jatz'katal* /jätz'-kab'-tal/ 'whip-world-come'), Mandarin *chen2-xi1* 'morning-sunlight,' Hani *aoq-bia bia* 'sky-bright/shining RED,' White Hmong *kaj-ntug* 'bright-sky,' Tetun *rai-naroma* 'land-to.grow.light,' and Vietnamese *bình minh* 'flat bright.' Moreover, Baruya, Buin, Badaga, Itzaj, Yana, Ancash Quechua, Hawaiian, and White Hmong colexify 'dawn' with '(day)light, bright' directly (ignoring glosses like 'first light of day'), and there are semianalyzable terms in Chukchi and Kaingang. The Itzaj association with an arrival by virtue of its term featuring a constituent meaning 'to come' is mirrored in Biloxi, which has *na^{n}´pi hu-di´* 'day come-??,' and such a term is also featured in Meyah. Similarly, Rotuman colexifies 'dawn' with 'come, arrive' directly, and Tsafiki has *oránan* /ora-nan/ 'good-come.closer.' Some languages have terms of the lexical type with a verb meaning 'to clear' or a noun 'clearness' as the second constituent, as in Buli *vari-nyaantiri* 'day clearness/brightness' and Muna *rara kamentae* 'clear morning.' Such a term is also found in Kiliwa, and Blackfoot and Aymara have semianalyzable terms with 'to clear, clear' as the meaning of the identifiable constituent (the Blackfoot term also denotes the 'clearing of weather'). Similarly, Rama has *sabítingi ngulaik* containing *sabítingi* 'clearing,' and Bakueri colexifies 'to dawn' with 'to clean;' note also the redundant Guaraní term *ko'e-t* 'dawn-clear.' Yir Yoront has *larr-mel-ngonngorr* 'day/place-eye-yesterday,' and 'dawn' is colexified with 'eye' directly in Koyraboro Senni among other meanings. Mali has *kunēnggunēng,* reduplicated from *kunēngga* 'sun, day,' Miskito *yu baiw-an* 'day sparkle-PAST.PTCPL,' Wichí *fwala'ihlo'*, which contains *fwala* 'day, sun' and the locative suffix *-lo'* 'in front of,' and terms betraying an association with 'sun' without colexification of 'day' are Badaga *ottu huṭṭu ~ hottu huṭṭu* 'sun be.born/rise,' Maxakalí *mãyõn xupep* 'sun arrive/leave'

and Manange *1tiŋi-2pʰja* 'sun-rise,' with a semianalyzable term in addition found in Aguaruna. There are also a number of other complex terms of the lexical type where one of the meanings is 'day.' Buli has *vayok* /vari-yok/ 'day-night,' Efik *ñkpö-usen* 'event-day,' Basque *egun-senti* 'day-feel,' Highland Chontal *egaŋwayda lidine jouba*, containing *lidine* 'day' and *jouba* 'finish,' Lesser Antillean Creole French *pwen di jou* 'point/fist/hand of day,' Itzaj *chun-k'in* 'base/trunk/foot-day' (which also means 'daytime, daylight'), Tuscarora *nyawę́ʻʔnuʻt*, containing *-ęT-* 'day' and *-aT-* 'stand' (alongside another similar term), Cavineña *huecaca japada-ma* 'day far-NEG,' and Hupda *wág hi-yćt* 'day FACTITIVE-lie.down.' Moreover, there are semianalyzable terms making reference to the 'day' in Buli, Kaluli, Sahu, Welsh, Nez Perce, Chayahuita, Embera, Piro, and Hawaiian. For the association with 'night' in the Buli and Kiliwa terms mentioned above, compare Buli *saliuk yok* 'morning nighttime' (which can also refer to the 'early morning'), Japanese *yoake*, analyzable as /yo-ake-Ø/ 'night-end-NR,' Cheyenne *vóone-ohtsé* 'all.night-go,' Piro *hoyetšno-kawa* 'night-period.of.time.following,' and that there is a term for 'dawn' which is derived from that for 'night' in Imbabura Quechua that can also mean 'early.'

There are some complex terms for 'dawn' where 'land' is the identifiable constituent: Pawnee has *huraahtaruuwispar*, containing *huraar* 'be land,' *ta-* 'suspended,' *iriwis* 'across,' and *war* 'walk,' and Tetun *rai-mutin* 'land-white' (compare also Burarra *ngana gunangarlcha*, containing *ngana* 'mouth' and *ngarlcha* 'become white,' Tehuelche *ʔoren ʔašk'en*, containing *ʔore ~ ʔoːre ~ ʔore* 'be white' and the locative nominalizer *-k'en*, as well as terms in various Romance languages going back to Latin *albus* 'white,' Buck 1949: 993), *rai-mutin* 'land-abundant,' and *rai-naroma* 'land-to.grow.light.'

As already seen from various examples discussed so far, 'dawn' is associated with 'morning' in a number of sampled languages. Khalkha has a derived term, Buli *saliuk yok* 'morning nighttime,' Mbum *rìm-péle* 'dark-morning,' Swahili *weupe wa alfajiri* 'witness of early.morning,' Muna *rara kamentae* 'clear morning,' Cayapa *dishquepenene* (/dishu-quepenene/), likewise analyzable as 'dark morning' (compare also the already mentioned Kiliwa *tiiy chip* 'dark/night clear/sweep' and that there is a semianalyzable term with 'dark' being the meaning of the identifiable constituent in Upper Chehalis and one featuring a verb meaning 'for darkness to disappear' in Cashinahua), Rama *tamas aik* 'morning side' and Fijian *mataka lailai* 'morning small' (this term, however, also means 'early morning' itself), Hawaiian *ʻehu kakahiaka* 'dust morning,' also figuratively denoting 'youth' and 'a shower that clears quickly,' and *moku ka pawa* 'be.cut DET darkness.before.dawn,' and Mandarin *chen2-xi1* 'morning sunlight.' The meanings 'dawn' and '(early) morning' are colexified in Buli (both by the term mentioned above and another one), Katcha, Yoruba, Muna, Badaga, Sora (by the term *'duŋroi ~ 'duŋroilen ~ 'duŋrojən*, presumably containing *duŋ-* 'get out of' and the continuative marker *-roi*), Cahuilla, Lake Miwok, Bororo, Cayapa, Embera, Macaguán, Toba (by the term *yo'oxoñi ~ yi'oxoñi*, containing *yi'oq* 'visible), and Fijian, and there are semianalyzable terms in Basque, Haida, Wintu, San Lucas Quiaviní Zapotec, Huambisa, and Fijian.

Gurindji, Badaga, Basque, Wintu, Aguaruna, Embera, Maxakalí, Hani, and Tetun colexify 'dawn' with 'sunrise' (an association present in some Indo-European languages as

well, Buck 1949: 993), sometimes by analyzable terms discussed either here or in section 155.

Moreover, Sora colexifies 'dawn' with 'cock-crow,' and Kwoma, more explicitly, has *apochoko wo nedii* 'cock crow time;' such an analyzable term is also featured in Piro. Hausa and Swahili colexify 'dawn' with 'morning prayer,' and Khoekhoe, Yoruba, Toaripi, and Samoan with 'twilight' (an association evidenced in Indo-European in various languages by a variety of structural types, Buck 1949: 993), Toaripi by a semianalyzable term containing an element meaning 'weather,' and Samoan by a term related to a verb meaning 'blaze, flare up' as said of a fire. Lake Miwok, Ancash Quechua, and Wayampi colexify 'dawn' with 'tomorrow' (Ancash Quechua also with 'to wake up' and Wayampi also with 'the whole night through, until dawn'). Southeastern Tasmanian, Toaripi, and Aguaruna colexify 'dawn' with 'dusk,' and in Efik and Buin, relevant terms are also greetings used in the morning. Similarly, Fijian *kida* also means "to go and salute a person on his arrival."

Other associations include: Welsh *gwawr* also means 'hue.' Muna *hawo-hawo rusa* is analyzable as 'return.home-RED deer' (since deer return to the woods at dawn), and *kowine* is also the name of a particular star or planet. Rotokas colexifies 'dawn' with 'pink' and Sentani *he-bo-* is analyzable as 'hang-knock/strike-.' Badaga *kari hakki ja:ma* is analyzable as 'black bird time,' Japanese *akebono* contains *ake* 'rise,' and Cheyenne *hosóvoomaeohtsé* and *hosóvoománo'e* contain *hosóvo* 'backward.' Kiliwa *(tiiy)-x-u?+saw-y* is analyzable as '(dark/night)-CAUS-OBL+see-ATT,' while Wintu colexifies 'dawn' with "spreading." Bororo has *baa aregodu* 'village appear,' Guaraní colexifies 'dawn' with 'year,' and Miskito has *lalma pauan* 'east reddened' and *lalma kahbi bara*, likewise containing *lalma* 'east' and *bara* 'when.' Hawaiian *ʻiao* also denotes the 'silversides' (a kind of small fish), the planet Jupiter as the morning star, and is a toponym for a site in West Maui, while *kaiao* also means 'to enlighten.' Lenakel colexifies 'dawn' and "a shine, something shining," and Malagasy *mangirandràtsy* is analyzable as /mangìrana-ràtsy/ 'to.have.chinks.through.which.the.light.shines-bad.' Samoan *vave-ao* is analyzable as 'early-dayflare,' and *tafa o ata* as 'be.visible LOC shadow.'

151. The Day

Representation: 94%
Motivated: 53.9%
Thereof Analyzable: 4.5% Thereof Colexifying: 49.4%
Thereof by Contiguity: 29.4% Thereof by Similarity: 0%
Recurrent associated meanings: sun, daylight/light, time, date, weather, noon, world, dawn, place/site, night in counting, occasion, hot/heat, week, watch/clock, morning, soil, afternoon, rain, sky, cloud, life, epoch/era

'Day' is very frequently colexified with 'sun' in the languages of the sample. This association is found in Hausa, Mbum, Rendille, Buin, Gurindji, Kyaka, Mali, Ngaanyatjarra, Sko, Toaripi, Japanese, Sora, Kildin Saami, Cahuilla, Ineseño Chumash, Comanche, Itzaj, Kiliwa (where the relevant term also means 'tawny, sun-colored'), Lake Miwok, Pawnee, Quileute, Xicotepec de Juárez Totonac, Yana, Yaqui, San Lucas Quiaviní Zapotec, Copainalá Zoque,

Lengua, Miskito, Rama, Tehuelche, Wichí, Fijian, Great Andamanese, Hawaiian, Manange, White Hmong, and Sedang. Furthermore, Dadibi has *giliga-de* 'sun-with' (in some of the sampled languages, there is also an association with 'moon' by virtue of the fact of it being colexified with 'sun,' compare section 60), Bororo *meri-ji* 'sun-DET,' and Tetun *loro-n* 'sun-INAL.POSS;' there are semianalyzable terms in Kwoma, Wappo, Kaingang, and Takia. By colexification of 'sun' and 'watch, clock' (compare section 79), the latter meaning is also colexified with 'day' in a small subset of these languages, namely Toaripi and Comanche. In Xicotepec de Juárez Totonac, the relevant term for 'sun, day' is derived from an element meaning 'warm,' while in Miskito and Hawaiian 'sun, day' is colexified with 'hot, heat' directly.

In Cahuilla, 'time' is colexified with 'sun' and 'day' in a single term, and the colexification of 'day' and 'time' alone is also attested in Efik, Ngambay, Kwoma, Kyaka (also with "event, opportunity, chance"), Meyah, Rotokas, Kosarek Yale, Yir Yoront, Sora, Carrier, Abipón, Aguaruna, Cavineña, Embera, Guaraní, Tehuelche, Great Andamanese, Lenakel, and Malagasy. Similarly, Yoruba, Basque, Khalkha, Highland Chontal, San Mateo del Mar Huave, Embera, Hawaiian, Samoan, and Tetun colexify 'day' with 'date' (San Mateo del Mar Huave also with 'name'). Furthermore, Jarawara has *yama wehe* 'thing/place/time light' (on *yama*, see § 6.4.3.15.), and Ngambay, Kyaka, and Yir Yoront colexify 'day' with 'place, site' and related meanings, such as 'soil' specifically in the latter two languages. Perhaps remotely similarly, Cubeo, Piro, Hawaiian, and Rotuman colexify 'day' with 'world' (Piro also with "time sphere" and 'expanse,' and Hawaiian also with 'to regain consciousness' inter alia).

As already suggested by the Jarawara term just mentioned, another recurrent association is that with '(day)light,' occurring by an analyzable term next to Jarawara in Efik (*uwem-ëyu* 'life-time/daylight;' compare colexification of 'day' and 'life' in Cayapa), and by colexification in Baruya, Buin (in which the relevant term *rua* also has the meaning 'door' due to accidental homonymy arising from borrowing), Kwoma, Kyaka, Sahu, Sentani, Yir Yoront, Badaga, Ket, Lesser Antillean Creole French, Nuuchahnulth, Guaraní, Hupda, Kaingang, Miskito, Yanomámi, Fijian, Hawaiian, Rotuman, and Samoan. The association is attested by semantic shift in Indo-European, as evidenced by cognates in Old Persian meaning 'day' and 'light' in Avestan, Greek, and Latin (Buck 1949: 991).

There are quite many cases in which 'day,' presumably more specifically in the sense of 'daytime,' is colexified with temporally contiguous phases of the day. Baruya, Wayampi, and Hawaiian colexify 'day' with 'dawn' (Yanomámi has *mɨ haru* 'face to.dawn' for 'daylight,' and Central Yup'ik has *erneq* /erte-neq/ 'dawn-thing.that.results.from'), Maxakalí with 'morning,' (and Aymara with the later hours of the morning specifically), Efik, Ngaanyatjarra, Sko, Wintu, and Yuki with 'noon,' Yoruba and Ket with 'afternoon' specifically, and Meyah, Aguaruna, Cubeo, Guaraní, and Malagasy with 'epoch, era.' The Yana, Yanomámi and Samoan terms also mean 'night,' but this meaning only occurs in counting time.

Yoruba, Kyaka, Haida, Itzaj, Oneida, Embera, Lenakel, and Malagasy colexify 'day' with 'weather' (Haida also with 'air,' and Embera also with 'atmosphere, climate'). Sentani and Bezhta colexify 'day' with 'rain,' Guaraní and Mandarin with 'sky,' and Hawaiian and

Samoan with 'cloud.' Kyaka, Sora, and Lenakel colexify 'day' with 'occasion.' Rotokas and Khalkha colexify 'day' with 'week' (Rotokas also with 'garden, work' and Khalkha also with 'planet').

Other associations include: Buli *vari*, as a verb, means 'to seize, take by force,' Hausa *yini* also denotes the "working part of the day" and the "withering of crops ... during the daytime" inter alia. Ngambay colexifies 'day' wth 'hour,' while Burarra *ngorrngurra* also means 'sleep' and is also the common name for the gecko inter alia (since geckos are held against the skin to soothe children so they fall asleep). Kyaka *gii* colexifies 'day' with 'season' and 'smile.' The Muna term *gholeo* also means 'dry season' as well as 'to dry in the sun' when used verbally, while Nunggubuyu *arara* also means 'to draw, to write.' Abzakh Adyghe *mafe* contains *f(e)* 'clear, bright,' while Bora *cójɨ̈ɨ́* also means 'early,' Wayampi *kɔ̃ʔɛ̃ ~ kɔʔɛ̃* also 'the whole night through, until dawn' and 'tomorrow,' Hani *nao* also 'to choke,' Bwe Karen *mu* also 'plant' inter alia, and *ni* also "woman's lower garment, skirt." Lenakel *nian* also means 'when,' and among the meanings of Hawaiian *lā* are also 'sail' and 'fin.' Samoan colexifies 'day' with 'celebration, party' and 'food for visitors.'

152. The Dusk

Representation: 55%
Motivated: 55.6%
Thereof Analyzable: 31.6% Thereof Colexifying: 24.8%
Thereof by Contiguity: 35.5% Thereof by Similarity: 8.1%
Recurrent associated meanings: dark/darkness, twilight, sunset, afternoon, evening, night, sun, dawn, small, recognize, bad

Most frequently, 'dusk' is associated lexically with 'dark,' 'darkness,' or 'to be dark.' By colexification, this is the case in Buli, Kosarek Yale, Khalkha (where one of the relevant terms colexifies 'clouded'), Welsh, and Lengua. As for complex terms, there is an interesting parallel between Lesser Antillean Creole French (*ti bwen* 'small dark') and Fijian (*buto-butō vaka-lailai* 'RED-dark DERIV-small). Other complex terms betraying the association with 'dark, darkness' are Efik *ë'kïm ñk'pö usen'ubök* 'darkness thing/event morning,' Basque *ilunabar* /ilun-nabar/ 'darkness-gray,' *ilun-alde* 'darkness-side/region/area/proximity,' and *ilun-senti* 'darkness-feel,' Upper Chehalis *kwɛtskwe´xta´n*, containing *kwɛts* 'middle' and *kwe´x* 'dark,' Chickasaw *oklhilisht ishtaya* containing *oklhilish* 'get dark' and *aya* 'go,' Kiliwa *tiiy nip* 'night/dark enter' and *?-nyaay=kw+tiiy* 'DN-sun/day=WH+dark/night,' Guaraní *ka'aru-pytû* 'afternoon-darkness,' Bwe Karen *khi-la* 'be.dark-down,' and Tetun *rai-nakaras* 'land-darken.' There are semianalyzable terms in Buin, Aymara, Jarawara, and Kaingang, and note also the similarity between Nez Perce *siẇwé·t* 'dusk' and *siẇé·t* 'dark,' as well as that between Comanche *tupɨsinawoni?* 'dusk' and *tupɨsibɨtɨ* 'dark color.' As the above discussion already makes clear, there are also associations with 'night' in some languages, sometimes due to colexification of 'night' and 'dark' (compare section 153). Alongside the associations in the abovementioned languages, and alongside direct colexification of 'dusk' and 'night' in Hani (by a semianalyzbale term containing *aoq* 'sky, heaven'), there are also many complex terms for the former where one of the constituents has the latter meaning. These are

Welsh *brig y nos*, containing *brig* 'top, summit' and *nos* 'night,' Highland Chontal *dihuama libuguih* /dihuamna libuguih/ 'walk night,' Quileute *ʔawí-sh* 'night-become,' Bororo *boeĉo paru* 'night beginning,' and Chayahuita *tashi-rin* 'night-CLASS.LARGE.AND.FLEXIBLE.' There are semianalyzable terms in Baruya, Copainalá Zoque, Jarawara, Kaingang, and Tehuelche, and moreover, Biloxi has *pûspûsi´~ pû´spûs ~ pĭspĭsi* which seems to be reduplicated from *psi ~ pŭs ~ pŭsi´* 'night.' The latter term colexifies 'dusk' with 'twilight,' an association also occurring (sometimes by one of the analyzable terms mentioned elsewhere in this chapter) in Efik, Khoekhoe, Rendille, Nunggubuyu, Toaripi (by a semianalyzable term featuring a constituent meaning 'weather' which also colexifies 'early dawn;' 'dawn' and 'dusk' are also not lexically distinguished in Southeastern Tasmanian and Aguaruna), Basque, Khalkha, Welsh, Wintu, and Hawaiian (by several terms, one of them being *mō-lehu* 'QUAL/STAT-ashes/ash.colored' which also means 'tipsy'), while Badaga has *sande jaːma* 'twilight time' (alongside another semianalyzable term containing *jaːma*). There are associations with '(early) evening' by colexification in Khoekhoe, Dongolese Nubian, Rendille, Upper Chehalis, and Quileute, and by analyzable terms in Yoruba (*werewere alẹ* 'quick evening'), Central Yup'ik (*ataku-ar(aq)* 'this.evening-little.piece.of'), and Samoan (*afiafi popogi* 'afternoon/evening for.the.night.to.fall'), and by a semianalyzable term in Buin. In turn, Baruya, Kwoma, Nunggubuyu, Basque, and Bislama colexify 'late afternoon' with 'evening,' and, as in Samoan, there are complex terms in Kaluli (*gaːlo nudáb* /gaːlo nudab/ 'afternoon taste'), Kwoma (*hogo ya yayi nedii* 'late.afternoon sun ladder time'), and Guaraní (*kaʼaru-pytû* 'afternoon-darkness'), as already mentioned. The Kwoma term just mentioned colexifies 'dusk' with 'sunset,' and this is also the case in Dongolese Nubian, Badaga, Basque, Bora, Huambisa, Ancash Quechua ('sunset with red color' specifically), and Tetun. Internal structure of the terms, if present, is discussed in section 156. However, many of them contain an element meaning 'sun.' Such terms for 'dusk' are Kwoma *hogo ya yayi nedii* 'late.afternoon sun ladder time,' Kyaka *neta anda penge dokopa* 'sun house departure when' and *neta anda penge gii* 'sun house departure time,' Kiliwa *ʔ-nyaay=kw+tiiy* 'DN-sun/day=PERF+dark/night,' Wayampi *kwalai-ɔ-ʔa-ɔ-ɔ* 'sun-3SG-fall-3SG-go' (colexifying 'west'), Hawaiian *liʻu-lā* 'slow-sun/day,' Sedang *hài pơxiammáng* 'sun/day begin/commence,' and Tetun *loro-teen* 'sun-excrement,' which also denotes a 'species of moss.' There are semianalyzable terms with an identifiable constituent meaning 'sun' in Kwoma, Cubeo and Huambisa.

Moreover, Kyaka has *yuu yasumi* 'time last.light.of.day' (and another semianalyzable term containing *yuu*) and Greek *likófōs* /lík-o-fōs/ 'wolf-STEM.FORMATIVE-light.' Interestingly, there are also three languages where terms for 'dusk' make reference to the fact that at dusk, it becomes hard to recognize the environment, and in particular other people. These languages are Muna (*dai wise* 'bad face'), Rotokas (*vuri evei* 'bad recognize'), and Sora (*erabmad'doja:lən* /er-ab-mad-'doja·l-ən/ 'NEG-CAUS-recognize-relative-N.SFX').

Although none of these meanings recurs exactly cross-linguistically, a number of the terms discussed above make reference to the fact that the sunset is a process by metaphorically using verbs of locomotion to convey the meaning, as in Chickasaw and Highland Chontal.

Other associations include: Buli *legi* also means 'startled, shocked,' Kyaka *yuu kwuange dokopa* contains *yuu*, meaning both 'earth, ground' and 'day' alongside still other meanings and *dokopa* 'when.' One *oi ninkle* is analyzable as 'bed garden,' and Abzakh Adyghe *pšepeze:χewey°e* as /pšape-ze:χewe-ɣ°e/ 'horizon-expand-time.' The Yuki term *hušámtik* contains *husám* 'day before yesterday,' and Cavineña *jaca-pude* is analyzable as 'leave-to.color.' Ancash Quechua *qarwayllu* appears to be related to *qarway* 'to ripen, acquire color,' while another Ancash Quechua term, *tsaqa*, also means 'spine.' Toba *napalñi* contains *apal* 'be opaque.' Tsafiki *pípuhuaqueno* contains *queno* 'to do,' and Great Andamanese *êrlōkorîtngarôi êr* 'place.' Kapingamarangi *heni* also means to "move around within an area," Hawaiian *mōlelehu* also 'drowsy, sleepy,' and *li'ulā* also means 'mirage, hallucination.'

153. The Night

Representation: 96%
Motivated: 26.7%
Thereof Analyzable: 7.0% Thereof Polysemous: 20.6%
Thereof by Contiguity: 24.6% Thereof by Similarity: 0%
Recurrent associated meanings: dark/get dark/darkness, evening, day in counting, black, sleep, thing

The most common association for 'night' is that with 'dark,' 'get dark,' and/or 'darkness' (present also in Indo-European evidenced by cognates in Sanskrit, Old Persian, and Ancient Greek, Buck 1949: 992). In the sample, the association is realized by colexification in Anggor, Berik, Buin, Kyaka (where the relevant term also denotes a kind of tree), Ngaanyatjarra (also with 'night sky'), Rotokas, Sahu, Badaga, Carrier, Upper Chehalis, Chickasaw, Highland Chontal, Kiliwa, Kiowa, Lesser Antillean Creole French, Yana (marked as dubious in the source), Yuki, Kaingang, Maxakalí, Miskito, Piro, Hawaiian, and Samoan. Ngambay has *lò-ndùl* 'time-black/dark,' Burarra *ana-munya* 'in-darkness,' Dadibi *huli-de* 'dark-who/with,' Kyaka *(yuu) iminjingi dokopa* '(earth) darkness when,' Bororo *boe ĉo* 'thing dark,' Jarawara *yama soki* 'thing be.black/dark' (compare § 6.4.3.15.), and White Hmong *tsaus-ntuj* 'dark-sky,' and derived terms, such as Ineseño Chumash *s-axiyi'* '3SG/3SG.POSS-be.dark' are found also in Nez Perce and Tuscarora; there is a semianalyzable term featuring a constituent meanind 'dark, black' in Lenakel. Furthermore, as for the association with 'black' in Ngambay and Jarawara, Buin, Kiowa, and Abipón colexifiy 'night' with 'black.'

'Night' is colexified with '(end of) evening' in Ngambay, Muna, Middle-Eastern and Southeastern Tasmanian, Abzakh Adyghe, Badaga, Welsh, Comanche, Lake Miwok, Wintu, Bwe Karen, and Yay (Yay also colexifies 'late afternoon'), mirrored in a connection by the common Indo-European root for 'night' with 'evening' in Hittite (Buck 1949: 992). Two languages of Australia, Ngaanyatjarra and Nunggubuyu, by a typical example of actual/potential-polysemy (O'Grady 1960), colexify 'night' with 'sleep' (the relevant Nunggubuyu term also denotes the "ant lion larva"). Terms for 'night' are also used as a

unit of counting time (as 'day' is used in English) in Rendille, Buin, Nunggubuyu, Haida, Yana, Yanomámi, Hawaiian (where this usage is obsolete), and Samoan.

Other associations include: Buli *yok* also denotes a "rafter of a thatched house," and Efik *okün'ëyu ~ okünoyu* appears to contain *ë'yu* 'sunshine, daylight.' Ngambay *tìl* also means 'shrub,' Buin *muu* also 'other,' Muna *alo* also 'dew, expose to dew,' Ngaanyatjarra *ngurra* also 'home, camp site, house, bed, bedroom,' and Sko *rángpang* contains *ráng* 'sun.' Toaripi *faita* is also the name of a type of red clay, and Yir Yoront *larr-ngonngorr* is analyzable as 'day/time-yesterday,' Badaga colexifies 'night' with 'twilight,' and is used adjectivally with the meaning "shadowy, shade" (another relevant term also means 'to be, exist,' and yet another one 'to cover, cover up'). Basque colexifies 'night' with 'to cover, close, shut, veil,' while Japanese *yoru* also means 'to twist' (the meanings are distinguished in writing). The Cahuilla term *túkmiyat ~ túkmaat* contains *-túk-* 'to go to bed, stay overnight.' Wintu *cipi* colexifies 'night' alongside 'evening' also with 'late,' and Huambisa *suwe* also means 'throat, chasm' (original Spanish gloss is 'garganta'). Bwe Karen *hɛ* also means 'to wander about' and 'to be hot' of food, and *nɛ* also denotes the spirit of a person. Hani *aoqqivq* also means 'dusk' (*aoq* is 'sky, heaven'), Hawaiian *pō* also denotes the "realm of the gods," "chaos, hell," and 'thick, dense,' as said of flowers and fragrance. Kapingamarangi *boo* is also used with the meaning "feeling of foreboding" in compounds and also denotes a "woman who is pregnant for the first time." Malagasy *àlina* is also a numeral for 'ten thousand.'

154. The Noon

Representation: 81%
Motivated: 65.8%
Thereof Analyzable: 58.3% Thereof Colexifying: 9.2%
Thereof by Contiguity: 51.7% Thereof by Similarity: 1.7%
Recurrent associated meanings: day, middle, sun, half, time, lunch, straight/right, afternoon, center, top, head, midnight, light/daylight, sky, hour, south, heart

Terms for the 'noon' are frequently analyzable of the lexical type, with one constituent being either 'day' or 'sun,' linked to the concept by contiguity. In a frequent type of terms, 'middle' is the meaning of the second constituent, as in Nuuchahnulth *ʔapẇin ṅaas* 'middle day' and Meyah *mówa ot déis* 'sun stand middle.' These are also very common in Indo-European (Buck 1949: 996). In the present sample, such terms are featured alongside Meyah and Nuuchahnulth in Efik, Hausa, Kanuri, Ngambay, Swahili, Kaluli, Sahu, Basque, Greek, Khalkha, Kildin Saami, Welsh, Blackfoot, Ineseño Chumash, Lake Miwok, Lakhota (where in addition a verb meaning 'to pass by' is present), Santiago Mexquititlan Otomí, Tuscarora, Central Yup'ik, Copainalá Zoque, Bora, Bororo (where in addition an element meaning 'sky' is present, compare also Cavineña *barepatya* /barepa-patya/ 'sky-middle'), Cubeo, Maxakalí (where the term is more specifically analyzable *mãyõn yãykote' yũm* /mãyõn yãykote yũm/ 'sun in.the.middle.of sit/be.located'), Ancash Quechua, Bislama, White Hmong, and presumably Haida, although this is not ultimately clear from the consulted source. Similarly, Upper Chehalis has *Pótûtsókûl*, containing constituents meaning

'middle' and 'set up.' Other similar terms, but without 'day' or 'sun' as contiguity anchor, are Buin *kugeniu-page* 'part.of.roof-middle' (compare Huambisa *tutupnirmatai*, presumably /tutupnik-matai/ 'straight-ridge/crest' and Malagasy *mitatào vovònana (ny àndro)* 'put.on.top ridgepole (?? day),' Cheyenne *-sétov-oéstá* '-middle-hang,' and Nez Perce *p̓éq-pe* 'be.middle-LOC' (for 'at noon'). 'Noon' and 'middle' are colexified directly in Piro. In this language, as well as in Sahu and Welsh, 'middle' is colexified with 'center' (in Piro also with 'part, interior'), and this association is mirrored in Mbum (*làú séséì* 'liver/center sun'); moreover, Bislama has *medel-dei* 'middle/center-day.' In Lake Miwok and Khalkha, 'middle' is colexified with 'half' (as well as with 'between' in Lake Miwok, and with 'midnight,' 'halfway, partly,' and 'mediocre, average' in Khalkha), and the terms in these languages therefore also betray an association with this meaning. They are not alone: Nivkh has *muɣv-n'lami* 'day-half,' a structure which is also found in Basque, Welsh, Kashaya (where in addition a verb meaning 'to break into two' is perhaps present, though the overall analysis of the term is unsure; note also that Nunggubuyu colexifies 'dawn' with 'to be or become torn or split'), and Imbabura Quechua. This pattern is also common in Indo-European (Buck 1949: 996). Similarly, Ket *suran* possibly consists of constituents meaning 'half' and 'daylight,' and there are also terms in other languages making reference to '(day)light' specifically rather than 'day,' namely Efik (*uwem-ëyu* 'life-time/daylight') and Manange (*1tʰoŋ-4sol* 'light-bright;' this term colexifies 'noon' with 'afternoon'). Returning to terms with 'sun' or 'day' as contiguity anchor, in three sampled languages, Burarra, Toaripi, and Pawnee, there is an association with 'top,' either by terms containing elements with that meaning directly (Burarra and Toaripi; in the latter language, an additional element meaning 'head' is present), or by a verb meaning 'to stand on top' (Pawnee). Moreover, Yir Yoront has *par-thila* 'head/top-hole:ERG/LOC.' The association with 'head' is also borne out in Sora (*teˑŋboːbˈjʊŋən* /teˑŋ-ˈbʔoːb-ˈjʊŋ-ən/ 'carry.on.head-head-sun-N.SFX'), Bororo (*barae etaiadaia ~ brae etaia keje*, containing *barae ~ brae* 'non-Indian' and *aia* 'middle of head,' literally therefore "the time when the sun is on top of the head of the white people"), Lengua (*yitsikso ikhim* /yitsiksik ikhim/ 'crown.of.head sun/day'), and Miskito (by the term *won lal kat*, containing *lal* 'head' and *kat* 'until'). There are also complex terms, often featuring a constituent meaning 'day,' that make reference to the fact that noon is lunchtime, such as Kwoma *a yadii nedii* 'eat day time.' Bezhta has *ƛobaƛ'as mex* 'noon.supper.gen time,' Kashaya *maʔa bumuyimeʔ*, containing *maʔa* 'food' and *meʔ* 'time,' while Ngaanyatjarra, Chickasaw, Highland Chontal, Haida, and San Lucas Quiaviní Zapotec colexify 'noon' with 'lunch' (see also Buck 1949: 996 for this association in Indo-European). For Huambisa *etsa tutupin* 'sun straight,' which was mentioned above, there are also parallels. Yoruba has *ọ̀sán-gangan* 'afternoon-right,' Yaqui *luula-katek-o* 'straight-be.sitting-COND,' Rama *kíbing kíngík* 'straight day,' and Tsafiki *yotú* /yo-tu/ 'sun-straight.' Similarly, Berik has *gwere bolap*, containing *gwer* 'sun' and *bolap* 'summit.' Ngambay has *kàrè wùr énje* 'sun/time/hour heart mother,' and Takia *ad bibe-n* 'sun heart-3SG.'

There are also other complex terms with 'sun' or 'day' acting as the contiguity anchor. These are Yoruba *agbede-meji ọjọ́* 'part-two day,' Muna *ghole-gholeo* 'RED-day,' Rotokas *ravireo vuuta* 'sun eventuate,' Toaripi *sare koko* 'sun/day narrow/contracted/restricted,' Itzaj *chumukk'in* /chu'm-Vk-k'in/ 'begin-DERIV-sun/day,' Kio-

wa *kʻiH-sɑ* 'day-AUG,' Yana *baloorpa*, containing the verb *ba-* 'for the sun to move, be in position' ("sun is southward up the mountain"), Aguaruna *étsa tajímai*, a literal translation of which would be "the sun is between the two sides," Toba *yi naʼaq*, perhaps 'richness day' (*yi* also has other meanings), Fijian *sigalevu tutu* /siga-levu tūtū/ 'day/sun-big stand,' Hawaiian *kau ka lā i ka lolo* 'place DET sun AGT DET brains,' Hani *aoq-nao naohhaol* 'sky-day daytime,' Bwe Karen *ləmu chitha* 'sun be.overhead,' Malagasy *antoandro be* /aN-to-àndro-be/ 'LOC-exact/true-day-big/much,' Samoan *tū-tonu o le lā* 'stand-be.exact GEN DET sun' and *ao-auli* 'day-be.pure,' as well as Tetun *loro-aas* 'sun-high' and *loro-natutu-n* 'sun-reach.its.peak-SINGULATIVE.' There is moreover a derived term in Carib, and Sko and Buli directly colexify 'sun' with 'noon,' while Efik, Ngaanyatjarra, Sko, Wintu, and Yuki colexify 'noon' and 'day.' There are semianalyzable terms where the identifiable constituent is 'sun' in Comanche, Aguaruna, Huambisa, and Rama, and where it is 'day' in Khoekhoe, Chukchi, San Mateo del Mar Huave, and Great Andamanese. Otherwise, the Cuisnahnuat dialect of Pipil has *wel-oːrah* 'before/well-hour' and Cayapa *catyu'ura* /catya-ura/ 'high hour' (this term colexifies, as do Khalkha and Cahuilla, 'midnight;' furthermore, due to colexification of 'sun' with 'time' and 'hour,' these associations are also present in a complex Ngambay term with the other constituent meaning 'middle'). Oneida colexifies 'noon' with 'south' (Wayampi has *yanɛ-alu-katu* 'our-south-good'), and Buli, Badaga, and Manange with '(early) afternoon.'

Other associations include: Abzakh Adyghe has *š(e)ǯaɣʷe*, containing *ǯ(e)* 'to study, call out' and *ɣʷe* 'time,' Badaga colexifies 'noon' with 'early morning' and 'off-hours' generally, while Khalkha *yde* is inter alia also the name of a river. Cahuilla *téklu-vel* is analyzable as 'stop-ABS,' and *máxeɬi-š is* derived from the verb *máxeɬi-* 'for the sun to be in the middle.' The Upper Chehalis term *ċúqʷ=alm* is analyzable as 'set.up=erect.object,' while Hupda *hi-mɨ̌ʔ-gʼet* is analyzable as 'FACTITIVE-under-stand.'

155. The Sunrise

Representation: 52%
Motivated: 65.6%
Thereof Analyzable: 52.3% Thereof Colexifying: 13.3%
Thereof by Contiguity: 16.5% Thereof by Similarity: 40.7%
Recurrent associated meanings: sun, come out/go out/emerge, dawn, appear, east, rise, light/daylight, day, sprout, arrive, split, be born, jump, burst out

Terms for the 'sunrise' are frequently morphologically complex and metaphor-based, with 'sun' (which sometimes is colexified with 'day') being the obvious contiguity anchor. Terms where the metaphor is based on notions such as 'to come out, go out, emerge,' such as Sora *duŋʲjʊŋən* /duŋ-jʊŋ-ən/ 'get.out.of-sun-N.SFX,' are also found in Hausa, Burarra, Meyah, Kolyma Yukaghir, Biloxi, Chickasaw, Itzaj, Kashaya, Lake Miwok, Ancash Quechua, and Hawaiian. Similarly, Nez Perce has *tiṅéhtit* /tiṅ-léht-it/ 'sun/moon-out,' Kaingang *rã vỹ jur mũ ha* 'sun TOP coming go now,' and 'sunrise' is colexified with a verb meaning 'to emerge, come out' in Dongolese Nubian, Upper Chehalis and Rotuman; in the latter two languages the additional sense 'burst out' is present inter alia. Another common association is that with the meaning 'to appear' (which often also means 'to rise' in the sampled

languages), such as Chukchi *tirk-inini* 'sun-appear.' Such terms are also found in Khoekhoe, Muna (by colexification, also with 'to come up,' 'to break open'), Khalkha, Cheyenne, Hawaiian, and Tetun. Tetun also has the alternative term *loron-mosu* 'day-to.appear' (there are semianalyzable terms containing an element meaning 'day' in Kyaka, Highland Chontal, Comanche, Chayahuita and Huambisa). The Tetun term colexifies 'east,' a pattern of colexification also found in Yoruba, Burarra, Kolyma Yukaghir, Nez Perce, and Bororo, in all cases by analyzable terms discussed elsewhere in this paragraph. Wintu has the term *puy ʔel-kawal*, seemingly analyzable as 'east inside-dawn.' Yana has *ʒauduloozi-,* which is, according to the source, "evidently from *ʒaudulauxauʒi* 'sun moves back out east-westward.'" 'Arrive' is also found in other languages which have complex terms of the lexical type as the meaning of the second constituent, as in Cubeo *aviái daino*, containing *aviá* 'sun' and *daino* 'arrival.' Such terms are also featured in Burarra, Kyaka (where an additional constituent meaning 'uphill, upwards' is present), and Maxakalí (where 'arrive' is actually colexified with 'leave'). Badaga has *ottu huṭṭu ~ hottu huṭṭu* 'sun/time be.born/rise,' and the 'birth'-metaphor is also attested in San Mateo del Mar Huave, which has *ajnchep teatnüt*, where *teatnüt* is 'sun' and *ajnchep* 'to be born, grow, bloom.' Alternatively, there is the term *ajnstop teatnüt*, with *ajnstop* meaning 'grow, bloom.' Similarly, Noni has *diuu ɛ san* 'sun to sprout,' and Khoekhoe *sore-s ǁhai-s* 'sun-3SG.FEM sprout/rise-3SG.FEM;' this association with 'sprouting' is also present in Hausa, where the verb meaning 'to come out' in the abovementioned term is colexified with 'to germinate,' as well as in Khalkha, also by colexification with another meaning, but in this case with 'to appear.' Kapingamarangi has *laa gu i hobo* 'sun DEC at jump,' and Samoan *oso aʻe* 'jump go.up/climb.' There are also terms making reference to the 'brightness' and 'light' that the rise of the sun brings. Rama has *núnik áungai* /núnik aungai/ 'sun light,' Hani *aoq-bia bia* 'sky-bright/shining RED,' Tetun *rai-naroma* 'day-to.grow.light,' while Cashinahua and Cayapa colexify 'sunrise' with 'light' or related meanings, and there is a semianalyzable term where the identifiable constituent is 'bright' in Sáliba. Yoruba has *ilà oòrùn* 'split sun,' and an analogous term is found in Efik and Noni.

In several of the above discussed languages, such as Khoekhoe, the second constituent is alongside the primary meaning also glossed with 'to rise' when occurring in the context of sunrise. However, there are also some terms, such as Efik *utïn asiakha* 'sun rise,' where the gloss of the complex term leaves it unclear whether the second constituent is only used in this context. Such cases are found also in Greek, Welsh, and Yuki. Genuine cases seem to be Nivkh *kʻjeṇ myrf yr* 'sun place.where.something.rises time,' Fijian *ni cadra na siga*, where *cadra* also means 'to rise' of seeds and the younger generation inter alia, and Lenakel, which colexifies "reach or arrive at a place higher than from which one started" with 'to rise,' said of the sun and the stars.

Other complex terms where one of the constituents is 'sun' are Kaluli *kowo:na: of-o:* 'little.lizard sun-??,' Ngaanyatjarra *tjirntukarrany(pa)*, containing *tjirntu* 'sun, day' and *karra* 'twilight' (this term colexifies 'sunrise' with 'sunset'), Toaripi *sare patai* 'sun/day ascend,' Nuuchahnulth *hiisakʷistuuɬʔitq hupaɬ* containing *kʷist* 'move away' and *hupaɬ* 'sun/moon,' Wappo *hìni čahwálseʔ* contains the word for 'sun, moon, clock, calendar' and a form of a verb meaning 'to crawl out.' Bororo has *meri ru-tu* 'sun fire-departure.' There are

semianalyzable terms where the identifiable constituent is 'sun' in Berik, Kaluli, Kyaka, Rotokas, Comanche, Guaraní, and Great Andamanese.

Gurindji, Badaga, Basque, Wintu, Aguaruna, Embera, Maxakalí, Hani, and Tetun colexify 'sunrise' with 'dawn,' Basque by the analyzable term *egun-senti* 'day-feel,' and Wintu by the analyzable term mentioned above.

Other associations include: Basque *goiz-argi*, also denoting the 'light of dawn,' is analyzable as 'morning-light,' and Upper Chehalis *pê´.tɛkχwoihwo´* contains the word for 'night.' Carrier *hae·aih en* is literally translated "where it comes from." Nuuchahnulth *huupkʷisɬaʔat* is analyzable as /hup-kʷisa-'ă'ʔa/ 'roundish.spherical.or.chunky.object-move.away-on.the.rocks,' while Pawnee *taʾa* is analyzable as /ta-a/ 'suspended-come.' Central Yup'ik *pit'e-* also means 'to take game' in the Yukon dialect, while Cayapa *dangueno* also means 'to shine.' Similarly, Tsafiki *chéino* and *cheyano* are related by unknown means to *chenu* 'illuminate.'

156. The Sunset

Representation: 55%
Motivated: 67.2%
Thereof Analyzable: 50.7% Thereof Colexifying: 17.7%
Thereof by Contiguity: 14.2% Thereof by Similarity: 45.1%
Recurrent associated meanings: sun, time, sink, enter, fall/drop, dusk, descend/go down/lower, day, disappear, die, west, twilight, dip, mountain, jump, sit down

As for 'sunrise,' of course, the meaning 'sunset' is expressed in many languages by complex terms where one of the constituents is 'sun' (which sometimes colexifies 'day'). These are often metaphorical in nature. Among them, the most common subpattern is the association with terms meaning 'to enter, entry,' as in Khoekhoe *sore-s ǂgâ-s* 'sun-3SG.FEM enter/set-3SG.FEM.' Such terms are also found in Ngambay, Yoruba, Comanche, Itzaj, and Cavineña, while in Nunggubuyu 'to enter' is colexified with 'for the sun to set;' the relevant term is in fact the reflexive of a verb meaning 'to put into.' Similarly, Rotokas has *ravireo rokore*, containing *ravireo* 'sun' and *roko* 'die out, go into, penetrate into,' Nez Perce *tiṅeyné·kt* /tiṅ-leyné·k-t/ 'sun/moon-into,' and San Mateo del Mar Huave *ajmel nüt* 'enter day' (for the association with 'day,' compare Kyaka *yuu kwualyamo*, containing *yuu* 'day, light' and *kwualyo* 'wipe, brush away, spread, shake out'). Otherwise, associations with verbs meaning 'to sink,' as in Kolyma Yukaghir *jel'o:d'əd-amlujbə* 'sun-sink,' are common (this term colexifies 'west,' which is also the case in the Bororo and Tetun complex terms already mentioned; Samoan has *tau-gāgaifo* 'go-a.little.towards.the.west,' and there is a semianalyzable term in Wintu which is also glossed as 'red sunset' and 'cloudy and sunshine'). Such complex terms also occur in Itzaj, Fijian, and Hawaiian, and the association is realized by colexification in Sora, Ineseño Chumash, Rotuman, and Samoan. Upper Chehalis *sʔísuʔs* contains *ʔísuʔ-* 'to dive' and a lexical affix 'face, eye, round object, dollar.' Meyah has *mówa esirí* 'sun fall,' and complex terms featuring verbs meaning 'to fall, fall in' or 'to drop' are also found in Hausa, Badaga (where an additional constituent meaning 'time' is present), Biloxi, Chickasaw, Itzaj, Bororo (where 'fall' is colexified with 'be born'),

Ancash Quechua (where 'to pounce' is colexified with 'to hide' and 'to bud'), and Tetun; moreover, the association is realized by direct colexification in Rendille, where the relevant term also means 'go down' and 'attack.' Cheyenne has *É-ta'ëhne* '3SG-disappear.' Complex terms based on a verb meaning 'to disappear' are also featured in Ngambay and Wappo, while Ineseño Chumash and Central Yup'ik colexify 'sunset' with 'disappear,' the former language also with 'to sink.' Wappo colexifies 'to disappear' with 'to die;' note that in Muna, *soo* is alongside other things a euphemistic expression for 'to die.' The Wappo association with 'die' is also mirrored in Katcha by colexification. Complex terms on the basis of meanings such as 'to descend,' 'to go down,' 'to lower,' such as Kwoma *ya yeyi nedii* 'sun go.down time' are found in Efik (alongside an element colexifying 'daylight' and 'time;' the entire term colexifies 'favor, privilege, opportunity'), Itzaj, Aguaruna, and Bislama (which also has *san i draon*, with *draon* meaning 'to drown' as well as 'for the sun to set,' inter alia); moreover, Buli and Rendille colexify the relevant meaning directly. Noni has *diuu ɛ sele* 'sun to jump,' and 'sunset' and 'fly, jump' are colexified in Lenakel. Wintu has *sasun pʰuyuq ʔolwiti harasin*, containing *sas* 'sun, moon,' *pʰuyuq* 'mountain,' *har* 'move' and *asin* 'away,' and a semianalyzable term containing a verb meaning 'to move' is attested in Yana (incidentally, 'hill, mountain' is among the meanings colexified with 'for the sun to set' in Rotuman). Kashaya has *ha·da cahci* 'sun sit.down,' and a term featuring a verb with the meaning 'sit down' alongside 'sun' is featured in Burarra as well. Greek and Welsh have terms containing elements meaning 'sun' and 'set.'

'Sunset' and 'dusk' are colexified in Dongolese Nubian, Kwoma (by the analyzable term *hogo ya yayi nedii* 'late.afternoon sun ladder time'), Badaga (by the analyzable term *sande jaːma* 'twilight time'), Basque (by the analyzable terms *ilun-alde* 'darkness-side/region/area/proximity/nearness' which itself colexifies 'twilight' and *ilunabar* /ilun-nabar/ 'darkness-gray'), Bora, Huambisa, Ancash Quechua (colexifying 'sunset' with 'red color' specifically), and Tetun (by a semianalyzable term containing *rai* 'land'), often also by analyzable terms discussed elsewhere in this paragraph, and Arabela has *nio nininiutianiyani* containing *nininiu* 'to be dusk;' note also the similarity between Xicotepec de Juárez Totonac *smalanka'nan* 'for the sun to set' and *smalanka'n* 'dusk' and Miskito *sáiwawa* 'sunset' and *sáiwan* 'dusk.' Efik colexifies 'sunset' with 'evening,' and Dongolese Nubian with 'early evening' specifically.

Other complex terms featuring a constituent meaning 'sun' are Kyaka *netame anda pelyamo/penge* containing *neta* 'sun,' *anda* 'house' and *penge* 'arrival' (there is another term featuring *neta* and *anda* as constituents the structure of which is not clear otherwise), Ngaanyatjarra *tjirntukarrany(pa)*, containing *tjirntu* 'sun, day' and *karra* 'twilight,' Toaripi *sare sukopai* /sare sukapai/ 'sun dip.into' (considered likely in the source; "[t]o the Elema, being a coastal people, the sun when setting seems to dip into the sea;" Kosarek Yale colexifies 'to dip into water' directly), Khalkha *nara(n) singge-ky* 'sun absorb/set-??,' Nivkh *kʻjeŋ uɣ yr* 'sun get.to mouth,' Itzaj *b'el k'in* 'sun go/travel/trip/going/exit,' Lesser Antillean Creole French *sòlei kouché* 'sun lie/sleep,' Aguaruna *tsawáut* /etsá-wáut/ 'sun-open.mouth,' Rama *núnik ausam altuang* containing *núnik* 'sun' and *altuang* 'waiting,' and Hani *naolma xavq li qavq* 'sun stay.overnight go PAST.' Further, there are semianalyzable terms where the identifiable constituent is 'sun' in Berik, Dadibi (where 'sun' and 'fire' are colexified),

Kaluli, Kyaka, Basque, Chukchi, Cubeo, Guaraní, Huambisa, Maxakalí, and Great Andamanese.

Other associations include: Ngaanyatjarra uses the same term for 'sunset' and 'sunrise,' Blackfoot *isttahkapi* also means 'to crawl in, under,' and *otahkoomaiksistoyi* contains *otahkoo* 'orange, yellow.' Carrier *na-e-ʻaih en* is analyzable as 'to.the.ground-??-get when,' whereas Upper Chehalis *stáqnč* contains *táqa-* cover/shade.' *Kwi´suɛnta´ne*, in contrast, contains *kwi´s* 'get dark, night.' Comanche *tabeʔikai ~ tabeʔikar+* also means 'evening,' while a literal translation of Kiliwa *ruwp ʔii* might be "shrouded from view." Pawnee *astaaruukita* contains *as-* 'foot' and *huukita* 'be on top' ("a metaphor, lit. the sun has its feet on (the horizon)"). Tuscarora colexifies 'sunset' with 'starset,' and Wintu has *puyel hololbeˑ*, containing *puyel* 'east hill' and *holol* 'sunshine, bright, light.' Chayahuita *i'hua-raya* is analyzable as 'a.short.time.ago-CLASS.FACE.EYE.OR.SEED.'

157. Man (Human Being)

Representation: 83%
Motivated: 35.1%
Thereof Analyzable: 8.9% Thereof Colexifying: 26.2%
Thereof by Contiguity: 7.2% Thereof by Similarity: 0.0%
Recurrent associated meanings: member of ethnic group, man, body, thing, owner/proprietor, mankind, woman, husband, earth/ground, child

Apart from use of the relevant terms as indefinite pronouns ('somebody'), which is disregarded here, terms with reference to 'human being, person' are commonly colexified with 'man' in the sense of 'male person' specifically, a pattern familiar from e.g. English (see Buck 1949: 79-80 for discussion of the broader Indo-European context). Precisely because of this, it is difficult to assess the strength of this pattern cross-linguistically, since the gloss 'man' is in fact ambiguous. Therefore, only cases are reported in which the glosses give reason to believe that 'male' specifically is genuinely colexified. This is the case in Efik, Kwoma, Yir Yoront, Basque, Chickasaw, Highland Chontal, San Lucas Quiaviní Zapotec, Cashinahua, Wayampi, and Bislama, while in Khoekhoe the meanings 'human being' and 'man, husband' are formed by using the same root, but suffixed with different nominal designants. Moreover, Sentani has the dvandva compound *do-mijɛ* 'man-woman,' and Takia *tamol-pein ~ tal-pein* 'man-woman' (this type of compound is common in New Guinea in general; for a precise parallel in the term for 'people' in Kalam see Pawley 1993: 99). Such a term is also found in Xicotepec de Juárez Totonac. Furthermore, Kanuri has *k-âm*, containing the prefix *k-* and the plural form of 'man,' and Guaraní has *yvy-póra* 'earth-dweller' (compare also Kaingang *ẽprã ke* 'on.ground make/say' and Latin *homo*, related to *humus* 'earth').

Another common association is the colexification of 'human, person' with a member of one's ethnic group in particular. This occurs in Koyraboro Senni, Burarra (by the term *gugaliya*, derived from *galiya ~ jaliya* 'to hear'), Gurindji, Nunggubuyu, Yir Yoront, Nivkh, Cahuilla, Haida (also by the redundant terms *xàaydlaa xàaydaraay* 'people visible.world/home,' which refers to the Canadian Haida specifically, and *xàayda giits'aads*

'people servants'), Lake Miwok, Quileute, Wintu, Yuki, San Lucas Quiaviní Zapotec, Bororo, Guaraní, Hupda, Miskito, Piro, Tsafiki, Yanomámi, Hawaiian, and Samoan. Furthermore, Efik, Ngaanyatjarra, Sentani, Yir Yoront, Copainalá Zoque, and Miskito colexify 'human being, person' with 'body' (Miskito also with 'flesh'), Yoruba, Badaga, and Hawaiian with 'mankind,' Kwoma and Bislama with 'husband,' and Ngambay, Bislama, and Hawaiian with 'owner, proprietor.' In Khoekhoe, the meanings 'man' and 'husband' are derived from the same root with different nominal designants (shift from 'person, human being' to 'male' and 'husband' is attested in Slavic according to Buck 1949: 80).

Kiowa has a general term 'human, person' identical segmentally to a verb meaning 'be alive' which takes additional suffixes depending on whether a man or a woman is referred to; a semianalyzable term probably featuring a constituent with that meaning is also found in Kiliwa and in Khalkha, where it colexifies 'animal.' Efik, Bororo, and Hawaiian colexify 'human, person' with 'thing' (Bororo also with 'time' and Hawaiian also with 'to say' inter alia), and in Swahili, the same root, associated with different prefixes, conveys these meanings. Koyraboro Senni *ibuna'adamayze* contains *ize* 'child;' the first constituent is borrowed from Arabic *'ibn 'aadam* 'son of Adam' (Hausa *'dan adan*, which has a similar structure is, when used in the singular, often applied to a person who has done something wrong), and Sahu and Hawaiian colexify 'person' and 'child' inter alia. Moreover, there is a semianalyzable term featuring a constituent meaning 'son, child' inter alia in Hani.

Other associations include: the formally redundant Ngambay term *kèje lè dèw* is analyzable as 'think GEN person' (*dèw* also means 'soul;' note that the Indo-European root giving rise to Engl. *man* etc. is, on one interpretation, connected with **men-* 'to think,' Buck 1949: 80). Buin *roi* also means 'gallbladder,' and Berik *angtane* also 'passenger' specifically. Kwoma colexifies 'human, person' with 'adult, mature,' 'front,' 'top,' 'finger, toe,' and 'twenty.' Muna *mie* is also used with reference to 'animals that build nests,' while Ngaanyatjarra *yarnangu* can also refer to the 'whole of something' and the 'appearance of something,' and Sentani *u* also means 'empty,' as of things. Tasmanian (Middle-Eastern) *kekána* contains *kána* 'language, utterance, speaking.' The Chukchi term *ʔorawetlʔan* contains *ʔoraŋ* 'openly' (the term is explained by Bogoraz 1922: 828 as quoted by Fortescue 2006: 269 as literally meaning 'one who walks openly,' which is used as a name for humans by malevolent spirits in Chukchi myth). Khalkha *kymyn* also means 'personality' and Lesser Antillean Creole French *moun* also means 'world, universe.' Blackfoot colexifies 'person' with 'pupil,' Itzaj *kristiyaanoj*, a loanword from Spanish, also denotes a 'Christian' specifically, and *mak ~ maak*, a native term, also means 'top.' Embera *ẽmberá* means 'person, personage' with masculine gender and 'people' with feminine gender. Guaraní *tekove* contains *teko* 'nature, character, being' and colexifies 'life' and other meanings. Huambisa *shuar* also means 'enemy,' and Hupda *húp* also 'good, new,' while Wayampi *tɛ-kɔ* is analyzable as 'NON.DETERMINATION-to.be.' Fijian *tamata* can also refer to chicken still in their eggs, and Hawaiian *kanaka* also means 'population,' 'pregnant,' and 'inhabited' inter alia.

158. The Saturday

Representation: 59%
Motivated: 33.9%
Thereof Analyzable: 33.8% Thereof Polysemous: 0.2%
Thereof by Contiguity: 22.9% Thereof by Similarity: 4.5%
Recurrent associated meanings: day, six, week, Saturn, Sunday, small, work, prepare, younger sibling, unique

Terms for 'Saturday' are overwhelmingly complex and of the lexical type in the languages of the sample, with 'day' (which sometimes is colexified with 'sun') being the meaning of one of the constituents. Apart from semianalyzable terms in Basque, Chayahuita, Guaraní, Fijian, and Manange, most frequently, 'six' is the meaning of the other constituent, as in Sko *bang nápánghì* 'day six' (*bang* is not explicitly glossed in the source, but since it recurs in the names of all other days of the week, it seems safe to assume that its meaning is 'day'). Such terms are also found in Ket, Chickasaw, Hani (where an additional constituent meaning 'week' is present; there is another semianalyzable term with this element), Hawaiian, Kapingamarangi, and Rotuman (where it is used by members of the Roman Catholic church). Similarly, Kaluli has *do:go:fe-yá* (/do:go:fe:-ya/) 'six-right.here' for 'on Saturday,' Mandarin *xing1qi1-liu4* 'week-six' and *li3-bai4-liu4* 'ritual-worship-six' (there are also other complex terms featuring a constituent with the meaning 'week': Swahili has *jumamosi* 'week first,' Kashaya *capaši· me?*, analyzable as /capa ši-w me?/ 'week make-ABS time,' 'Saturday' in Oneida is derived from a term meaning 'daylight, week,' and 'Saturday' and 'week' are colexified directly in Central Yup'ik by the analyzable term *maqi-neq* 'steambath-thing.that.results.from'). Ngambay has *ndɔ kùlà mìsán* 'day work six,' and Kiliwa *ti?chat=msirlhpaayp* 'work=six' (note also the rare Hawaiian term *lā hoʻomalolo* 'day cease.work'). In contrast, Vietnamese has *ngày thứ bảy* 'day ordinal seven,' the variation presumably due to differences in which day is the day of rest and prayer and differences in which day is taken to be the first of the week.

Japanese has *do-yō* 'earth/Saturn-day' and Welsh *dydd Sadwrn* 'day Saturn.' 'Saturday' is colexified with 'Saturn' in Badaga and Khalkha (see § 6.4.3.13.6. for the history of this pattern, as well as Buck 1949: 1007-1009 on the coexistence of the 'ecclesiastical' system and the 'planetary' system in Indo-European). Abzakh Adyghe and Basque have *mefezaqᵊʔe* and *egu-bakoitz* respectively, both analyzable as 'day-unique.'

Rotokas has *Topekakau voki* 'tobacco day' ("comes from the old practice of paying wages with tobacco on Saturday;" compare also the fact that an Upper Chehalis term for 'Saturday' contains a word for 'time' and a verb meaning 'to distribute,' with the explanation in the source offered having recourse to the fact that on Saturday rations were distributed). A literal translation of Blackfoot *to'tohtáátoyiiksistsiko* according to the consulted source is "the day before the holy day," that is, Sunday, Chickasaw *Nittak Hollo' Nakfish* contains *nittak* 'day' and *nakfish* 'younger sibling of the same sex,' Tuscarora has *awę́·ʔnakwt*, literally 'day alongside,' Fijian *Siga Vakarau leka*, containing *siga* 'day' and *leka* 'small,' Samoan *Aso Toʻonaʻi* 'day collect.food.in.preparation.for.Sunday.meal' (Lenakel has *(nian taha) n-epinapine-aan* /(nian taha) n-epinapina-aan/ '(day BENEFACTIVE) NMLZ-prepare-

NMLZ,' a term which is restricted to the speech of older Christians, and Rotuman *(teràn) a'itạ* /(teraṇi) a'itạ/ '(day) prepare;' this term is used by Wesleyan Methodists). All the denominations just mentioned, starting with that in Blackfoot, make reference in some way to the fact that 'Saturday' is contiguous temporally to 'Sunday,' the most important day of the week. The Fijian association with 'small' is explained by the fact that the 'Saturday' is a "small" holiday when compared with the Sunday, and this interpretation is corroborated by terms such as Cheyenne *Tšêške'-ma'heóneéšeëva* 'little-Sunday,' Kiowa *dā'k'ɪH-syḥn-gyH* 'Sunday-small-NOUN.POSTFIX,' and the fact that Carrier has a semianalyzable term where the identifiable constituent is 'small.' Moreover, Biloxi has *noxwi´so^{n}´tka* /noxwi´di so^{n}tka´da/ 'Sunday younger.brother,' and there is a semianalyzable terms involving a word for 'Sunday' in Haida ('Sunday' in Haida is *s@ndii*. Haida also features the term *sandiigaa cajuu*, where *cajuu* is 'small,' and *sandiigaa* is also suspiciously similar to Engl. *Sunday*. The source notes that this term is "[s]aid to derive from the missionary habit of flying a small flag on Saturday and a big one on Sunday," which suggests that this is a term of the Cheyenne and Kiowa type just mentioned) and Nez Perce (where the literal translation offered in the source is "toward Sundaying"). Moreover, Kyaka has *koro kuki* 'period.of.time small' (this term also denotes a 'pause').

Other associations include: Buli *Asibi* is also a name given to children born on a Saturday, and Badaga *cani ~ sani*, colexifying 'Saturn,' also means 'bad omen, ill fate' and similar things. Pawnee *Piiriku?* contains *iirik* 'to see,' the literal translation and explanation given in the source being "seeing different ones (when people went into town on Saturdays to shop)."

159. The Virgin

Representation: 38%
Motivated: 53.3%
Thereof Analyzable: 53.0% Thereof Colexifying: 2.7%
Thereof by Contiguity: 36.0% Thereof by Similarity: 7.1%
Recurrent associated meanings: girl/daughter, unmarried girl/woman, man, NEG, closed/shut, young, female/woman, husband

Terms for 'virgin' are in some languages of the sample analyzable, containing an element meaning 'man' or 'husband' and a negator alongside other morphemes. For one, Ngambay has *gère ngàw àáng* 'know man/husband NEG,' and similar terms, with verbs meaning 'to feel' and 'to see' rather than 'know' respectively are found in Chickasaw and Bora. Kyaka has *wanakeme akali nyii range*, containing *wanake* 'girl,' *akali* 'man, husband,' *nyii* 'take!' and *range* 'self, ego' (alongside other semianalyzable terms featuring a constituent meaning 'man, husband),' and Piro *makloji jeji metkatowa*, containing *makloji* 'girl,' *jeji* 'man,' and *metkatowa*, which seems to be related to *metkatu* 'blind.' Ngaanyatjarra *wati-ku ngurrpa* is analyzable as 'man-of ignorant.' Embera has *w'ẽ̌ra aw'ẽ̃rakirú,* containing *w'ẽ̌ra* 'woman' and *aw'ẽ̃rakirú*, an adjective meaning 'new, virgin.' Fijian *goneyalewa savasavā* is analyzable as 'girl clean,' and *dau lato* seems to contain *dau* 'to commit adultery' (*lato* is the name of a tree species).

There are derived term from roots meaning 'unmarried' in Khoekhoe and Lenakel ('virgin' is colexified with 'unmarried girl/woman' in Kwoma, Kyaka, Nez Perce, Pawnee, Central Yup'ik and Lenakel). Samoan colexifies 'female' generally with 'virgin,' semianalyzable terms where the identifiable constituent is 'woman' are featured in Kwoma and Guaraní, and Tetun has *feto-raan* 'woman-bleed,' referring to menstruation. This is also the case in Upper Chehalis (*máymáy̓sx̣amn̓* contains *máya-* 'enter' and *x̣am-* 'menstrual period). A term referring to the physical rather than cultural aspects of virginity is also found in Yanomámi (*ka kõmi* 'opening closed'). Yir Yoront and Hawaiian have similar terms of the lexical type, and Efik a derived term from a verb meaning 'be closed.'

Haida and Miskito colexify '(be) virgin' with '(be) young' (Japanese has the term *otome*, questionably analyzable as /oto-me/ 'young-woman'), while in Muna, Khalkha, Welsh, Nez Perce, Pawnee, Samoan, relevant terms can also refer to a '(teenage) girl,' 'daughter,' and sometimes even 'child' more generally.

Other associations include: Noni *wan wvu tfu* contains *wan* 'child' and *tfu* 'each,' Baruya *kwaiyagaala muja'* makes reference to the fact that it is forbidden to address a virgin in the male speech (*yagaala*) concerning marriage, and Burarra *(-)yawuk* also denotes a 'childless girl or woman.' The Muna term *bungasa* also means 'untouched,' as said e.g. of palm trees not yet tapped and *kalambe*, as a verb, is glossed as 'become a girl,' while Sahu *mosolese* also denotes the "the wife of a geneaologically younger member of a hereditary group." Kashaya *cap^{h}ya* also means 'celibate.' Oneida *yáh thaʔtewaʔalyá·ku* contains the negative particle *yáh* and the lexical roots *-aʔal-* 'net, lace,' and *.yaʔk-* "detach, sever, break, cut in two." San Lucas Quiaviní Zapotec *digài'* is also a name for the 'five centavo coin,' Wintu *lo·ymes* is derived from *lo·y* 'put on front apron, become an adolescent girl,' 'virgin spirit, faery,' while Chayahuita *nanon miáchin* contains *nanon* 'girl' and *miáchin*, which means 'a bit, rather' when following an adjective or adverb and 'much' otherwise. Great Andamanese *ôtlêkinga* also means 'poor,' while Hawaiian *puʻu-paʻa* is analyzable as 'mound-firm' and also denotes the 'kidneys' and, figuratively, 'emotions, affections.'

160. The Widow

Representation: 80%
Motivated: 32.1%
Thereof Analyzable: 26.9% Thereof Colexifying: 5.2%
Thereof by Contiguity: 22.3% Thereof by Similarity: 1.5%
Recurrent associated meanings: woman, husband, die/dead, orphan, unmarried, NEG, single, prostitute, lose, grieve/mourn, alone

Many of the motivated terms for this meaning are complex and of the lexical type, with one constituent being 'woman,' and the other indicating the differentia specifica. A common subtype is that making reference to the fact that a widow has 'lost' her husband, that is, that he has 'died,' as for instance Hawaiian *wahine kāne make* 'woman husband die,' Tsafiki *puyamin sono*, presumably containing *puyano* 'to die' and *sono* 'woman,' or Chickasaw *ihoo hattak imilli-'* 'woman man lose-NMLZ.' Such terms are also featured in Efik, Katcha, Lesser Antillean Creole French, and Samoan. This particular denomination strategy need

not always be realized by terms containing a lexical element meaning 'woman:' next to derived terms in Carib and Manange, Kanuri has *kám-bà* 'person/man-not,' Itzaj *ix-kimen-'icham* 'FEM.NOUN.CLASS-dead-husband' and *ix-ma'-'icham* 'FEM.NOUN.CLASS-NEG-husband,' Kiliwa *?kuswa?=p-i?+hiw* 'husband=MP-DIST+fly,' Lake Miwok *míiw-helak* 'husband-lack,' and Cavineña *eahue-maju-que* 'husband-die-who.' Similarly, Baruya has *kwaimaaya'* /kwalai-maaya-da' ~ kwalai-maaya-sa'/ 'man-without-she' and Rotokas *oira asava*, containing *asa* 'without.' The association with 'dying' is also found in Japanese (*mi-bō-jin* 'not.yet-die-person'), Cahuilla (*múk-vel* 'get.sick.or.weak/die-ABS.NMLZ'), and there are semianalyzable term where the identifiable constituents are verbs meaning 'to die' in Piro and Lenakel.

An association present also a few times in Indo-European (Buck 1949: 131) is that with the meaning 'alone' (as is the case in Old Norse and some of its daughters, Buck 1949: 131): Pawnee has *capaktihuks*, literally 'alone woman' (which is also quite literally used for a woman who lives alone), Bororo *aredu koadureudo* 'woman alone/widowed,' Burarra colexifies the relevant meanings, and there are semianalyzable terms where the identifiable constituent is 'alone' in Blackfoot and Yanomámi. In the latter language, the relevant term colexifies 'widow' with 'deprived of a loved one or goods,' 'with empty hands,' and 'orphan;' colexification with 'orphan,' as in one case in Indo-European according to Buck (1949: 131), is also encountered in Rendille, inter alia in Abzakh Adyghe, Chickasaw, and Wintu, where the relevant term *lolcit* also may refer to anyone in mourning, a 'survivor,' and someone possessed by the ghost of the deceased husband or wife, and hence 'cursed;' indeed, it is related to the root *lol* 'bereaved, ghost of dead spouse or divorced spouse; orphan(ed), phantom.' Similarly, Rotokas has *virakoiva*, with the gender/number marker *–va* 'feminine singular' for 'widow' and *virakoito* with *–to* 'masculine singular' for 'orphan,' a case of grammatical alternation. Toaripi has *ua lelesi* 'woman unmarried,' denoting "a woman without a husband (for any reason)" (which is also said of the relevant term in Swahili) and 'widow' and 'unmarried (woman)' are colexified in Efik, Buin, and Fijian. Khoekhoe has *!oa-tara-s* 'grieve-woman-3SG.FEM,' and Biloxi *a´xti a^{n}tcodo^{n}* /a´xti´ a^{n}-tcodo^{n}/ 'woman mourn.' There are also other complex terms with one constituent being 'woman': Mbum has *wű̃ yìù* 'woman celibate' (this term also sometimes refers to a "prostitute or widow who accept young boys to sleep with," 'widow' and 'prostitute' are also colexified in Efik) and *wű̃ kúl* 'woman single' (compare colexification of 'single' and 'widow' in Welsh), Kwoma has *mibiya mima*, where *mima* is 'woman' and *mibiya* refers to the state of having been married before but not anymore, due to death of the spouse or divorce, Kyaka has *enda waiya (petenge)* 'woman chopped (living),' *enda mee peta-mo doko* 'woman empty sticking-ASSOC that' (this term has the variant *etembo peta-mo* 'single sticking-ASSOC'), Kosarek Yale *youwok kelabo* /youwi-ok kelabo/ 'sterile-just/alone women,' Chukchi *janraŋaw* /janra-ŋew/ 'separate-woman' (note that the root **weidh-* 'separate' is the origin of some Indo-European terms for 'widow,' Buck 1949: 131), Nivkh *tyrmu-umgu* 'lonely woman,' Cheyenne *otôxaa'é'e* perhaps literally 'uncovered woman,' San Mateo del Mar Huave *lemben omal najtaj*, seemingly containing *omal* 'point' alongside *najtaj* 'woman,' Santiago Mexquititlan Otomí *'ra-nxu* 'one-woman,' Hani *miqcyuq* /almiq-cyuq/ 'woman-poor,' and White Hmong *poj-ntsuam* 'woman-destitute.one.' In addition, there are semianalyzable

terms with a constituent meaning ‘woman’ and/or ‘wife’ in Berik, Dadibi, Japanese, Highland Chontal, Cubeo, Miskito, and Great Andamanese.

Other associations include: Buli *pokong* is also the name of a particular tree and its fruit, Hausa *bazawara ~ zawara* also denotes “anything which has been used and is to be re-used” inter alia, while Noni *kpwɛɛ ŋkfu* appears to contain *ŋkfu* ‘late.’ Buin *rarupere* also means ‘banana garden,’ Burarra colexifies ‘widow’ with “divorced or lone person,” Meyah *óna mesina* ‘widower’ is analyzable as ‘male string.bag,’ and Nunggubuyu *nᵍayi* also denotes “any close kin of dead person” generally. Muna *kowaluno* ‘widow in mourning period’ is related to *walu* ‘to shroud’ and *kowalu* ‘shroud,’ since in Muna culture widows are wearing a white sarong over their heads in the 100 day mourning period they undergo. Ngaanyatjarra *wanakaarla* may be related to *wana* ‘digging stick’ (compare Dixon 2002: 99 for this association in Australian languages), Yir Yoront *thum-kuwn(l)* appars to be analyzable as ‘fire-grandchild,’ and *yoq-warry+pann* is analyzable as ‘tree-poor.fellow+younger.sibling.’ Abzakh Adyghe *pxʼəż* is analyzable as /pxʼə-żə/ ‘girl-old,’ and Badaga *kuṇḍe* colexifes ‘widow’ with ‘woman’ generally (though this is archaic), and *muṇḍe* also means ‘bad woman, bad wife.’ Ineseño Chumash *ʼunitaxiš* contains *ʼunitax-* ‘to leave behind,’ while Wappo *ƙà· hanchóya*, containing *ƙá·* ‘person,’ also may refer to a “lonely person, person without friends or family, poor person.” Central Yup’ik *aipaineq* is analyzable as /aipaq-(ng)ite-neq/ ‘partner/mate/spouse-lack-thing.that.results.from.’ Huambisa colexifies ‘widow’ with ‘sister-in-law/brother-in-law,’ while Fijian *dawai* also denotes “one who has none to care for him and her,” and Malagasy *mpitòndra tèna* is literally ‘one who carries oneself.’

References

Abbi, Anvita. 2006a. Endangered languages of the Andaman Islands. Munich: Lincom Europa.

Abbi, Anvita. 2006b. Vanishing voices: a typological sketch of Great Andamanese. In: Anju Saxena and Lars Borin (eds.): Lesser-known languages of South Asia. Status and policies, case studies and applications of information technology, 107-123. Berlin: De Gruyter.

Abbott, Clifford. 2000. Oneida. Munich: Lincom Europa.

Adam, Lucien. 1968. Langue Mosquito. Grammaire, vocabulaire, textes. Nendeln: Kraus.

Adelaar, K.A. 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Oceanic Linguistics 28, 1-46.

Adelaar, Alexander. 2009. Malagasy vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/28.

Adelaar, Willem F.H. with Pieter C. Muysken. 2004. The languages of the Andes. Cambridge: Cambridge University Press.

Aguirre Licht, Daniel. 1999. Embera. Munich: Lincom Europa.

Aikhenvald, Alexandra Y. 1999. Areal diffusion and language contact in the Içana-Vaupés Basin, North West Amazonia. In: Dixon and Aikhenvald (eds.), 385-415.

Aikhenvald, Alexandra Y. 2000. Classifiers: a typology of noun classification devices. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2003. A grammar of Tariana, from Northwest Amazonia. Cambridge: Cambridge University Press.

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In: Alexandra Y. Aikhenvald and R.M.W. Dixon (eds.): Serial verb constructions: a cross-linguistic typology, 1-68. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2007. Typological distinctions in word-formation. In: Shopen (ed.), 1-65.

Aikhenvald, Alexandra Y. 2008. The Manambu language of East Sepik, Papua New Guinea. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y., and R.M.W. Dixon. 2001, eds. Areal diffusion and genetic inheritance: problems in comparative linguistics. Oxford: Oxford University Press.

Alinei, Mario. 1983, ed. Atlas linguarum Europae. Volume 1, commentaires: 1er fasc. Assen: Van Gorcum.

Alinei, Mario. 1996. Theoretical aspects of lexical motivation. In: Maj Reinhammar (ed.): Mål i sikte. Studier i dialektologi tillägnade Lennart Elmevik, 1-10. Uppsala: Almquist & Wiksell.

Alinei, Mario. 1997. Principi di teoria motivazionale (iconimia) e di lessicologia motivazionale (iconomastica). In: Luisa Mucciante and Tullio Telmon (eds.): Lessicologia e lessicografia. Atti del XX Convegno della Societa Italiana di Glottologia (Chieti-Pescara, 12-14 ottobre 1995), 11-36. Rome: Il Calamo.

Alinei, Mario. 2001. Aspects of a theory of motivation (iconymy). Versus. Quaderni di studi semiotici 88/89, 89-97.

Allwood, Jens. 2003. Meaning potentials and context: some consequences for the analysis of variation in meaning. In: Hubert Cuyckens, René Dirven, and John R. Taylor (eds.): Cognitive approaches to lexical semantics, 29-65. Berlin: De Gruyter.

Alpher, Barry. 1991. Yir-Yoront lexicon. Sketch and dictionary of an Australian language. Berlin: De Gruyter.

Alpher, Barry J. and David G. Nash. 1999. Lexical replacement and cognate equilibrium in Australia. Australian Journal of Linguistics 19, 5-56.

Alves, Mark. 2009. Vietnamese vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/24

Ameka, Felik K., and Stephen C. Levinson 2007. Introduction: the typology and semantics of locative predication: posturals, positionals and other beasts. Special issue of Linguistics 45, 847-871.

Anderson, Gregory D.S., and K. David Harrison. 2008. Sora. In: Gregory D.S. Anderson (ed.): The Munda languages, 299-380. London: Routledge.

Anderson, John M. 1991. Notional Grammar and the redundancy of syntax. Studies in Language 15, 301-333.

Anderson, John M. 2011a. The substance of language. Vol. 1: the domain of syntax. Oxford: Oxford University Press.

Anderson, John M. 2011b. The substance of language. Vol. 3: phonology-syntax analogies. Oxford: Oxford University Press.

Anderson, Ralph Dewitt. 1963. A grammar of Laz. Dissertation, University of Texas at Austin.

Anderson, Stephen R. 1985. Typological distinctions in word formation. In: Shopen (ed.), 3-56.

Andrade, Manuel J. 1933. Quileute. New York: Columbia University Press.

Anonymous. n.d. Organised phonological data. Buin (Telei, Terei Language) [BUO]. Manuscript, Summer Institute of Linguistics.

Aoki, Haruo. 1970. Nez Perce grammar. Dissertation, University of California at Berkeley.

Aoki, Haruo. 1994. Nez Perce dictionary. Berkeley/Los Angeles: University of California Press.

Applegate, Richard B. Ineseño Chumash grammar. 1972. Dissertation, University of California at Berkeley.

Apresjan, Juri D. 1974. Regular polysemy. Linguistics 142, 5-32.

Apresjan, Yuri D. 1992. Lexical semantics. User's guide to contemporary Russian vocabulary. Ann Arbor: Karoma.
Archivo de Lenguas Indígenas de México. 1983. Huave de San Mateo del Mar, Oaxaca. San Ángel: Centro de Investigación para la Integración Social.
Arcodia, Giorgio F. 2007. Chinese: a language of compound words? In: Fabio Montermini, Gilles Boyé, and Nabil Nathout (eds.): selected proceedings of the 5th Décembrettes: morphology in Toulouse, 79-90. Somerville: Cascadilla Proceedings Project.
Armbruster, Charles Hubert. 1960. Dongolese Nubian: a grammar. Cambridge: Cambridge: University Press.
Armbruster, Charles Hubert. 1965. Dongolese Nubian: a lexicon. Nubian-English, English-Nubian. Cambridge: Cambridge University Press.
Arroyo, Víctor Manuel. 1955. Elementos de gramática Otomí. Patrimonio Indígena del Valle del Mezquital.
Auroux, Sylvain, E.F.K. Koerner, Hans-Josef Niederehe, and Kees Versteegh. 2000, eds. History of the language sciences. Berlin: De Gruyter.
Austin, P., R. Ellis, and L. Hercus. 1976. 'Fruit of the eyes'. Semantic diffusion in the Lakes languages of South Australia. Papers in Australian Linguistics 10, 57-77.
Awóbùlúyì, Ọládélé. 1992. Aspects of contemporary Standard Yoruba in dialectological perspective. In: Akinwumi Iṣọla (ed.): New findings in Yoruba studies, 1-79. Lagos: J.F Ọdunjọ Memorial Lectures, University of Ibadan.
Baayen, R. Harald. 2009. languageR: data sets and functions with 'analyzing linguistic data: a practical introduction to statistics'. R package version 0.955.
Bach, Adolf. 1969. Deutsche Mundartforschung. Ihre Wege, Ergebnisse, und Aufgaben. Third Edition. Heidelberg: Winter.
Bai, Hosabi, and Clyde M. Whitby. 2008. Dadibi-Tok Pisin-English: po dage dabe. Ukarumpa: Summer Institute of Linguistics.
Bakker, Dik. 2010. Language sampling. In: Song (ed.), 100-127.
Bakker, Dik, and Ewald Hekking. 2009. Otomi vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/35
Balakrishnan, Ramaswami. 1999. Badaga: a Dravidian language. Annamalainagar: Annamalai University.
Barcelona, Antonio. 2000a, ed. Metaphor and metonymy at the crossroads: a cognitive perspective. Berlin: De Gruyter.
Barcelona, Antonio. 2000b. On the plausibility of claiming a metonymic motivation for conceptual metaphor. In: Barcelona (ed.), 31-58.
Barcelona, Antonio. 2002. Clarifying and applying the notions of metaphor and metonymy within Cognitive Linguistics: an update. In: Dirven and Pörings (eds.), 207-277.
Bargery, G.P., and Diedrich Westermann. 1934. A Hausa-English dictionary and English-Hausa vocabulary. Oxford: Oxford University Press.
Barlaan, Rodolfo R. 2003. Animism exploits linguistic phenomena. In: Eugene H. Casad and Gary B. Palmer (eds.): Cognitive Linguistics and non-Indo-European Languages, 173-192. Berlin: De Gruyter.

Barnes, R. H. 1977. *Mata* in Austronesia. Oceania 48, 300-319.
Barsalou, Lawrence W. 1992. Frames, concepts, and conceptual fields. In: Lehrer and Kittay (eds.), 21-74.
Bastian, Adolf. 1870. Sprachvergleichende Studien. Leizig: F.A. Brockhaus.
Bastian, Adolf. 1881. Der Völkergedanke im Aufbau einer Wissenschaft vom Menschen. Berlin: Dümmler.
Bates, Douglas, and Martin Maechler. 2009. lme4: linear mixed-effects models using S4 classes. R package version 0.999375-32.
Bauer, Laurie. 1983. English word-formation. Cambridge: Cambridge University Press.
Bauer, Laurie. 2002. What you can do with derivational morphology. In: Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer and Maria D.Voeikova (eds.): Morphology 2000: selected papers from the 9th Morphology Meeting, Vienna, 24-28 February 2000, 37-48. Amsterdam: Benjamins.
Bauer, Laurie. 2009. Typology of compounds. In: Lieber and Štekauer (eds.), 343-356.
Bauer, Robert S. 1992. Winkin', blinkin', and nod: a study in historical-comparative semantics of Southeast Asian languages. Linguistics of the Tibeto-Burman Area 15, 151-184.
Beare, John Isaac. 1908. The works of Aristotle. Transl. into English under the editorship of W. D. Ross. Vol 3: De parva naturalia: de sensu et sensibili, de memoria et reminiscentia, de somno, de somniis, de divinatione per somnum. Oxford: Oxford University Press.
Beck, David. 2005. The lexicalization of morphosyntactically complex expressions. Paper presented at the 9th International Cognitive Linguistics Conference, Seoul.
Becker, Alton L. 1975. A linguistic image of nature: the Burmese numerative classifier system. Linguistics 165, 109-121.
Becker, Thomas. 2002. Autohyponymy: implicature in lexical semantics, word formation, a nd grammar. Journal of Germanic Linguistics 14, 105-136.
Beckwith, Christopher I. 2009. Empires of the Silk Road. A history of Central Eurasia from the Bronze Age to the present. Princeton: Princeton University Press.
Beekes, Robert. 2010. Etymological dictionary of Greek, Vol. 2. Leiden/Boston: Brill.
Behrens, Leila. 2000. Semantics and typology. Sprachtypologie und Universalienforschung 53, 21-38.
Behrens, Leila. 2002. Structuring of word meaning II: aspects of polysemy. In: Cruse et al. (eds.), 319-337.
Behrens, Leila, and Hans-Jürgen Sasse. 1997. Lexical typology: a programmatic sketch. Arbeitspapier Nr. 30 (Neue Folge). Cologne: Institute for Linguistics, University of Cologne.
Benaissa, Taik. 1991. Vocabulario Sáliba-Español, Español-Sáliba. Lomalinda/Meta: Editorial Alberto Lleras Camargo.
Bender, Ralf, and Stefan Lange. 2001. Adjusting for multiple testing - when and how? Journal of Clinical Epidemiology 54, 343-349.
Benveniste, Émile. 1935. Origines de la formation des noms en Indo-Européen. Paris: A. Maisonneuve.

Bergin, Thomas Goddard, and Max Harold Fisch. 1984. The New Science of Giambattista Vico. Ithaca: Cornell University Press. [Translation of Vico 1744/1976]

Bergsland, Knut. 1986. Comparative Eskimo-Aleut phonology and lexicon. Journal de la Société Finno-Ougrienne 80, 63-137.

Berlin, Brent. 1972. Speculations on the growth of ethnobotanical nomenclature. Language in Society 1, 51-86.

Berlin, Brent. 1992. Ethnobiological classification. Princeton: Princeton University Press.

Berndt, Catherine H., and Ronald M. Berndt. 1951. An Oenpelli monologue: culture-contact. Oceania 22, 24-52.

Betz, Werner. 1949. Deutsch und Lateinisch. Die Lehnbildungen der Althochdeutschen Benediktinerregel. Bonn: Bouvier.

Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. Language 79, 708-736.

Bickel, Balthasar 2007. Typology in the 21st century: major current developments. Linguistic Typology 11, 239-251.

Bickel, Balthasar. 2008. A general method for the statistical evaluation of typological distributions. Manuscript, University of Leipzig.

Bickel, Balthasar, and Johanna Nichols. 2009. The geography of case. In: Andrej Malchukov and Andrew Spencer (eds.): The Oxford Handbook of Case, 479-493. Oxford: Oxford University Press.

Bickel, Balthasar, and Johanna Nichols. In Press. Oceania, the Pacific Rim, and the theory of linguistic areas. Berkeley Linguistics Society 32.

Biggs, Bruce. 1965. Direct and indirect inheritance in Rotuman. Lingua 13, 383-415.

Biggs, Bruce. 1978. The history of Polynesian phonology. In: Stephen A. Wurm and Lois Carrington (eds.): Second international conference on Austronesian linguistics: proceedings, 691-716. Canberra: Australian National University.

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: counting and beyond. In: Jadranka Gvozdanovic (ed.): Numeral types and changes worldwide, 113-185. Berlin: De Gruyter.

Bisang, Walter. 2006. Linguistic areas, language contact and typology: some implications from the case of Ethiopia as a linguistic area. In: Yaron Matras, April McMahon, and Nigel Vincent (eds.): Linguistic areas: convergence in historical and typological Perspective, 75-98. Hampshire: Palgrave MacMillan.

Blackings, Mairi, and Nigel Fabb. 2003. A grammar of Ma'di. Berlin: De Gruyter.

Blank, Andreas. 1997. Prinzipien des Lexikalischen Bedeutungswandels am Beispiel der Romanischen Sprachen. Tübingen: Niemeyer.

Blank, Andreas. 2001. Pathways of lexicalization. In: Haspelmath et al. (eds.), 1596-1608.

Blank, Andreas. 2003. Words and concepts in time: towards diachronic cognitive onomasiology. In: Regine Eckardt, Klaus von Heusinger and Christoph Schwarze (eds.): Words in time. Diachronic semantics from different points of view, 37-65. Berlin: De Gruyter.

Blevins, Juliette, and Andrew Wedel. 2009. Inhibited sound change. An evolutionary approach to lexical competition. Diachronica 26, 143-183.

Bloomfield. Leonard. 1933/1984. Language. New York: Holt.

Blust, Robert A. 2005. Liver and lungs: a semantic dyad in Austronesian languages. Oceanic Linguistics 44, 537-543.

Blust, Robert. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. Phonology 24, 1-36.

Blust, Robert A. 2009. The Austronesian languages. Canberra: Pacific Linguistics.

Boas, Franz. 1911. Handbook of the American Indian languages. Washington: Government Printing Office.

Boeder, Winfried. 2005. The South Caucasian languages. Lingua 115, 5-89.

Bogoraz, Vladimir G. 1922. Chukchee. In: Franz Boas (ed.): Handbook of American Indian Languages, U.S. Buerau of American Ethnology, Bulletin 40.2. Washington: Bureau of American Ethnology.

Bohnemeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke, and Felix K. Ameka. 2007. Principles of event segmentation: the case of motion events. Language 83, 495-532.

Bolker, Benjamin M., Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M. Henry H. Stevens, and Jada-Simone S. White. 2008. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution, 127-135.

Bopp, Franz. 1836. Vocalismus oder sprachvergleichende Kritiken über J. Grimm's deutsche Grammatik und Graff's althochdeutschen Sprachschatz mit Begründung einer neuen Theorie des Ablauts. Berlin: Nicolaische Buchhandlung.

Bowden, John. 1992. Behind the preposition. Grammaticalisation of locatives in Oceanic languages. Canberra: Australian National University.

Bowden, Ross. 1997. A dictionary of Kwoma: a Papuan language of North-east New Guinea. Canberra: Australian National University.

Bradley, David. 2002. Some distinctive characteristics of the Burmic lexicon. In: Cruse et al. (eds.), 1070-1071.

Bredin, Hugh. 1984. Metonymy. Poetics Today 5, 45-58.

Brentano, Franz. 1956. Die Lehre vom richtigen Urteil: nach den Vorlesungen über Logik mit Benützung anderer Manuskripte aus dem Bereich der Erkenntnistheorie aus dem Nachlaß. Ed. by Franziska Mayer-Hillebrand. Bern: Francke.

Bright, William. 1960. Animals of acculturation in the California Indian Languages. University of California Publications in Linguistics 4, 215-246.

Bright, William. 2004. 'Analyzability' of nouns in Northwestern California. Paper presented at the Conference on the Ecology of Language, University of California, Berkeley.

Bright, Jane O., and William Bright. 1965. Semantic structures in Northwest California and the Sapir-Whorf hypothesis. American Anthropologist 67, 249-258.

Brinton, Laurel J. 2000. The structure of modern English: a linguistic introduction. Amsterdam: Benjamins.

Brinton, Laurel J., and Elizabeth Closs Traugott. 2005. Lexicalization and language change. Cambridge: Cambridge University Press.
Britton, A. Scott. 2005. Guaraní concise dictionary. New York: Hippocrene Books.
Brown, Bob, and Walsa Translation Team. 2007. Diksenari Walsana moa, Pisinna moa, Englisna moa = A short dictionary of the Walsa (Waris) language, Tok Pisin and English. Online: http://www.sil.org/pacific/png/abstract.asp?id=48988.
Brown, Cecil H. 1983. Where do cardinal direction terms come from? Anthropological Linguistics 25, 121-161.
Brown, Cecil H. 1984. Language and living things. Uniformities in folk classification and naming. New Brunswick: Rutgers University Press.
Brown, Cecil H. 1999. Lexical acculturation in Native American languages. Oxford: Oxford University Press.
Brown, Cecil H. 2001. Lexical typology from an anthropological point of view. In: Haspelmath et al. (eds.), 1178-1190.
Brown, Cecil H. 2005a. Finger and hand. In: Haspelmath et al. (eds.), 526-529.
Brown, Cecil H. 2005b. Hand and arm. In: Haspelmath et al. (eds.), 522-525.
Brown, Cecil H., and Stanley R. Witkowski. 1981. Figurative language in a universalist perspective. American Ethnologist 8, 596-615.
Brown, Cecil H., and Stanley R. Witkowski. 1983. Polysemy, lexical change and cultural importance. Man 18, 72-89.
Brown, H. A. 1968. A dictionary of Toaripi with English-Toaripi index. Sydney: University of Sydney.
Brown, H. A. 1972. The Elema languages. A comparative study of Toaripi, Orokolo and their related dialects. Dissertation, University of London.
Brown, Herbert A. 1977. The Eleman language family. In: Karl Franklin (ed.): The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea, 281-376. Canberra: Pacific Linguistics.
Brugmann, Karl. 1900/1981. Über das Wesen der sogenannten Wortzusammensetzung. Eine sprachpsychologische Studie. In Leonhard Lipka and Hartmut Günther (eds.): Wortbildung, 135-178. Darmstadt: Wissenschaftliche Buchgesellschaft. Originally published in Berichte über die Verhandlungen der königlich-sächsischen Gesellschaft der Wissenschaften zu Leipzig, Philologisch-historische Classe 52, 359-401.
Brumbaugh, Robert. 1987. The Rainbow Serpent on the Upper Sepik. Anthropos 87, 25-33.
Buchler, Ira R., and Kenneth Maddock. 1978, eds. The Rainbow Serpent: a chromatic piece. The Hague: Mouton.
Buck, Carl Darling. 1929. Words for world, earth and land, sun. Language 5, 215-227.
Buck, Carl Darling. 1949. A dictionary of selected synonyms in the principal Indo-European languages. Chicago: Chicago University Press.
Buckwalter, Alberto S. 1980. Vocabulario Toba. Buenos Aires: Talleres Gráficos Grancharoff.

Buenaventura, Edgar V. 1993. Observaciones preliminares acerca del idioma Macaguán: a puntes culturales, fonología, apuntes gramaticales, voabulário Macaguán – Español. Santafé de Bogotá: Instituto Lingüístico de Verano.

Burenhult, Niclas. 1996. Deep linguistic prehistory with particular reference to Andamanese. Lund Working Papers in Linguistics 45, 5-24.

Burenhult, Niclas. 2008a, ed. Language and landscape: geographical ontology in cross-linguistic perspective Special issue of Language Sciences.

Burenhult, Niclas. 2008b. Streams of words: hydrological lexicon in Jahai. Language Sciences 30, 182-199.

Burenhult, Niclas, and Stephen C. Levinson. 2008. Language and landscape: a cross-linguistic perspective. Language Sciences 30, 135-150.

Bybee, Joan. 1988. The diachronic dimension in explanation. In: John A. Hawkins (ed.): Explaining language universals, 350-379. Oxford: Blackwell.

Bybee, Joan L. 1995. Diachronic and typological properties of morphology and their implications for representation. In: Laurie B. Feldman (ed.): Morphological aspects of language processing, 225-246. Hillsdale: Lawrence Erlbaum.

Callaghan, Catherine A. 1963. A grammar of the Lake Miwok language. Dissertation, University of California at Berkeley.

Callaghan, Catherine A. 1964. Phonemic borrowing in Lake Miwok. In: William Bright (ed.): Studies in Californian linguistics, 46-53. Berkeley/Los Angeles: University of California Press.

Callaghan, Catherine A. 1965. Lake Miwok dictionary. Berkeley/Los Angeles: University of California Press.

Callaghan, Catherine A. 1979. Miwok color terms. International Journal of American Linguistics 45, 1-4.

Camac, Mary K., and Sam Glucksberg. 1984. Metaphors do not use associations between concepts, they are used to create them. Journal of Psycholinguistic Research 13, 443-455.

Camargo, Eliane. 1991. Phonologie, morphologie et syntaxe. Étude descriptive de la langue Caxinauá (Pano). Dissertation, Paris.

Camden, William G. 1979. Parallels in structure of lexicon and syntax between New Hebrides Bislama and the South Santo language as spoken at Tangoa. Papers in Pidgin and Creole Linguistics 2, 51-117.

Camp, Elizabeth L., and Millicent R. Liccardi. 1989. Diccionario Cavineña-Castellano, Castellano-Cavineña, con bosquejo de la gramática Cavineña. Dallas: Summer Institute of Linguistics.

Campbell, Lyle. 1975. Constraints on sound change. In: Karl-Hampus Dahlstedt (ed.): The Nordic languages and modern linguistics 2, 388-406. Stockholm: Almquist & Wiksell.

Campbell, Lyle. 1979. Middle American languages. In: Lyle Campbell and Marianne Mithun (eds.): The languages of Native America: historical and comparative assessment, 902-1000. Austin: University of Texas Press.

Campbell, Lyle. 1985. The Pipil language of El Salvador. Berlin: De Gruyter.

Campbell, Lyle. 1996. On sound change and challenges to regularity. In: Durie and Ross (eds.), 72-89.
Campbell, Lyle. 1998. Historical Linguistics. An introduction. Edinburgh: University of Edinburgh Press.
Campbell, Lyle. 2009. Areal linguistics. In: Keith Brown and Sarah Ogilvie (eds.): Concise encyclopedia of languages of the world, 62-68. Oxford: Elsevier.
Campbell, Lyle, and Jon Ringen. 1981. Teleology and the explanation of sound change. In: Wolfgang U. Dressler, Oskar E. Pfeiffer, and John R. Rennison: Phonologica, 57-68. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. Language 62, 530-570.
Campbell, Lyle, and Mauricio J. Mixco. 2007. A glossary of Historical Linguistics. Edinburgh: Edinburgh University Press.
Campbell, Robin N., and Theresa Bowe Macdonald. 1983. Text and context in early language comprehension. In: Margaret Donaldson, Robert Grieve, and Chris Pratt (eds.): Early childhood development and education. Readings in psychology, 115-126. Oxford: Blackwell.
Canart, Paul. 1979. Studies in comparative semantics. New York: St. Martin's Press.
Canzio, Riccardo. 1997. Text and music in Bororo chanting. Société Suisse des Américanistes / Schweizerische Amerikanisten-Gesellschaft Bulletin 61, 63-70.
Capell, A. 1991. A new Fijian Dictionary. Suva: Government Printer.
Carneiro, Robert L. 2003. Evolutionism in Cultural Anthropology. A critical history. Boulder, Co.: Westview Press.
Carranza Romero, Francisco, and Wolf Lustig. 2003. Diccionario Quechua Ancashino-Castellano. Madrid: Iberoamericana / Frankfurt: Vervuert.
Carrington, Lawrence D. 1984. St. Lucian Creole: a descriptive analysis of its phonology and morpho-syntax. Hamburg: Buske.
Casagrande, Joseph B. 1954. Comanche linguistic acculturation II. International Journal of American Linguistics 20, 217-237.
Cassidy, Frederic G. 1971. Tracing the Pidgin element in Jamaican Creole (with notes on method and the nature of Pidgin vocabularies). In: Hymes (ed.), 203-221.
Cengerova, Zuzana, Yen-Ling Chen, Roey Gafter, Patrick Hall, Sverre Johnsen, E. J. Kim, Ricardo Lezama, Pamela Munro, Hyejin Nah, Bruno Olsson, Olivia N. Sammons, Matthias Urban, and Oliviana Zakaria. 2009. Notes on Kashaya as spoken by Anita Silva. LSA Institute, Berkeley.
Chaffin, Roger. 1992. The concept of a semantic relation. In: Lehrer and Kittay (eds.), 253-288.
Chamberlain, A.F. 1890. The Thunder-Bird amongst the Algonkins. American Anthropologist 3, 51-54.
Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralization. Phonetica 50, 28-43.
Charles-Luce, Jan. 1997. Cognitive factors involved in preserving a phonemic contrast. Language and Speech 40, 229-248.

Chomsky, Noam. 1975. Reflections on language. New York: Pantheon.
Chowning, Ann. 1996. POc **mata*: How many words, how many meanings? In: John Lynch and Fa'afo Pat (eds.): Oceanic studies: proceedings of the first International Conference on Oceanic Linguistics, 47-60. Canberra: Pacific Linguistics.
Charney, Jean Ormsbee. 1993. A grammar of Comanche. Lincoln: University of Nebraska Press.
Churchward, C. Maxwell. 1998/1940. Rotuman English Dictionary of 1940. Suva: Institute of Pacific Studies, University of the South Pacific.
Citarrella, Cinzia. 2010. Review of Vanhove (2008). Linguist List 21.3510.
Clark, Eve V. 1981. Lexical innovations: how children learn to create new words. In: Werner Deutsch (ed.): The child's construction of language, 299-328. London: Academic Press.
Clark, Eve V. 1982. The young word-maker: a case study of innovation in the child's lexicon. In: Eric Wanner and Lila R. Gleitman (eds.): Language acquisition: the state of the art, 390-425. Cambridge: Cambridge University Press.
Clark, Eve V. 2000. Coining new words: old and new word forms for new meanings. In: Lise Menn and Nan Bernstein Ratner (eds.): Methods for studying language production, 53-67. Mahwah, NJ: Lawrence Erlbaum.
Clark, Eve V., and Barbara Frant Hecht. 1982. Learning to coin agent and instrument nouns. Cognition 12, 1-24.
Clements, G. N. 2003. Feature economy in sound systems. Phonology 20, 287-333.
Coates, William Ames. 1968. Near-homonymy as a factor in language change. Language 44, 467-479.
Cole, Peter. 1982. Imbabura Quechua. Amsterdam: North Holland.
Collins, Allan M., and M. Ross Quillian. 1969. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior 8, 240-247.
Colman, Fran, and John Anderson. 2004. On metonymy as word-formation: with special reference to Old English. English Studies 85, 547-565.
Comenius, Johann Amos. 2005/1648. Novissima linguarum methodus. Ed. by Gilles Bibeau, Jean Caravolas, and Claire Le Brun-Gouvanic. Geneva: Droz.
Comrie, Bernard. 1993. Typology and reconstruction. In: Charles Jones (ed.): Historical Linguistics: problems and perspectives, 74-97. London: Longman.
Comrie, Bernard. 2000. Language contact, lexical borrowing, and semantic fields. In: Dicky Gilbers, John Nerbonne, and Jos Schaeken (eds.): Languages in contact, 73-86. Amsterdam: Rodopi.
Comrie, Bernard. 2005. Alignment of case marking of pronouns. In: Haspelmath et al. (eds.), 398-401; 404-405.
Comrie, Bernard. 2008. Linguistic diversity in the Caucasus. Annual Review of Anthropology 37, 131-143.
Comrie, Bernard, and Sandra A. Thompson. 2007. Lexical nominalization. In: Shopen (ed.), 334-381.
Comrie, Bernard, and Madzhid Khalilov. 2009a. Bezhta vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/15.

Comrie, Bernard, and Madzhid Khalilov. 2009b. Loanwords in Bezhta, a Nakh-Daghestanian language of the North Caucasus. In: Haspelmath and Tadmor (eds.), 414-429.
Conzemius, Eduard. 1929. Notes on the Miskito and Sumu languages of Eastern Nicaragua and Honduras. International Journal of American Linguistics 5, 57-115.
Corbera Mori, Angel Humberto. 1994. Fonologia e gramática do Aguaruna (Jívaro). Dissertation, Universidad de Campinas.
Corbett, Greville G. 1991. Gender. Cambridge: Cambridge University Press.
Corbett, Greville G. 2007. Gender and noun classes. In: Shopen (ed.), 241-279.
Costello, Fintan J., and Mark T. Keane. 2000. Efficient creativity: constraint-guided conceptual combination. Cognitive Science 24, 299-349.
Costello, Fintan J., and Mark T. Keane. 2001. Testing two theories of conceptual combination: alignment versus diagnosticity in the comprehension and production of combined concepts. Journal of Experimental Psychology: Learning, Memory, and Cognition 27, 255-271.
Cotari, Daniel 1978. Diccionario Aymara-Castellano, Castellano-Aymara. Cochabamba: Instituto de Idiomas, Padres de Maryknoll.
Courtz, Henk. 2008. A Carib grammar and dictionary. Toronto: Magoria Books.
Cowan, H. K.J. 1965. Grammar of the Sentani language, with specimen texts and dictionary. 'S Gravenhage: Martinus Nijhoff.
Craig, C. G. 1990. A grammar of Rama. Lyon: Université de Lyon.
Cram, David, and Jaap Maat. 2000. Universal language schemes in the 17th century. In: Auroux et al. (eds.), 1030-1043.
Cristofaro, Sonia. 2005. Purpose clauses. In: Haspelmath et al. (eds.), 506-509.
Croft, William. 1993. The role of domains in the interpretation of metaphors and metonymies. Cognitive Linguistics 4, 335-370.
Croft, William. 2000. Explaining language change: an evolutionary approach. Harlow: Longman.
Croft, William. 2003. Typology and universals. Second Edition. Cambridge: Cambridge University Press.
Croft, William, and D. Alan Cruse. 2004. Cognitive Linguistics. Cambridge: Cambridge University Press.
Crosswhite, Katherine. 1999. Inter-paradigmatic homophony avoidance in two dialects of Slavic. UCLA Working Papers in Linguistics 1, Papers in Phonology 2, 1-18.
Crowell, Thomas Harris. 1979. A grammar of Bororo. Dissertation, Cornell University.
Crowley, Terry. 2003. A new Bislama dictionary. Second Edition. Suva: Institute of Pacific Studies.
Crowley, Terry. 2004. Bislama reference grammar. Honolulu: University of Hawai'i Press.
Cruse, D. Alan. 1986. Lexical semantics. Cambridge: Cambridge University Press.
Cruse, D. Alan. 2002. Hyponymy and its varieties. In: Rebecca Green, Carol A. Bean, and Sung H. Myaeng (eds.): The semantics of relationships: an interdisciplinary perspective, 3-22. Dordrecht: Kluwer.

Cruse, D. Alan, Franz Hundsnurscher, Michael Job, and Peter Rolf Lutzeier. 2002, eds. Lexikologie/Lexicology. Berlin: De Gruyter.
Cumont, Franz. 1935. Les noms des planètes et l'astrolatrie chez les Grecs. L'Antiquité classique 4, 5-43.
Curnow, Timothy Jowan. 2001. What language features can be 'borrowed'? In: Aikhenvald and Dixon (eds.), 412-436.
Curnow, Timothy Jowan, and Anthony J. Liddicoat. 1998. The Barbacoan languages of Colombia and Ecuador. Anthropological Linguistics 40, 384-408.
Cysouw, Michael. 2002. Interpreting typological clusters. Linguistic Typology 6, 69-93.
Cysouw, Michael. 2003. Against implicational universals. Linguistic Typology 7, 89-101.
Cysouw, Michael. 2005. Quantitative methods in typology. In: Gabriel Altmann, Reinhard Köhler, and Rajmond G. Piotrowski (eds.): Quantitative linguistics: an international handbook, 554-578. Berlin: De Gruyter.
Cysouw, Michael. 2010. Dealing with diversity: towards an explanation of NP-internal word order frequencies. Linguistic Typology 14, 253-286.
Cysouw, Michael, and Bernard Comrie. Forthcoming. Some observations on typological features of hunter-gatherer languages.
Dahl, Östen. 2005. Tea. In: Haspelmath et al. (eds.), 554-557.
Dahl, Östen. 2008. An exercise in a-posteriori language sampling. Sprachtypologie und Universalienforschung 61, 208-220.
Dahl, Östen, and Viveka Velupillai. 2005. The past tense. In: Haspelmath et al. (eds.), 268-2 69; 276-277.
Davidson, Matthew. 2002. Studies in southern Wakashan (Nootkan) grammar. Dissertation, State University of New York at Buffalo.
Dayley, Jon P. 1989a. Tümpisa (Panamint) Shoshone grammar. Berkeley/Los Angeles: University of California Press.
Dayley, Jon P. 1989b. Tümpisa (Panamint) Shoshone dictionary. Berkeley/Los Angeles: University of California Press.
De Haan, Ferdinand. 2005. Semantic distinctions of evidentiality. In: Haspelmath et al. (eds.), 314-317.
Derbyshire, Desmond C., and Doris L. Payne. 1990. Noun classification systems of Amazonian languages. In: Doris L. Payne (ed.): Amazonian linguistics: studies in lowland South American languages, 243-71. Austin: University of Texas Press.
Dez, Jacques. 1980. Structures de la langue malgache: eléments de grammaire à l'usage des francophones. Paris: Publications Orientalistes de France.
Dickinson, Connie. 2002. Complex predicates in Tsafiki. Dissertation, University of Oregon.
Diez, Friedrich Christian. 1875. Romanische Wortschöpfung. Bonn: Weber.
Dirven, René. 1985. Metaphor as a basic means for extending the lexicon. In: Wolfang Paprotté and René Dirven (eds.): The ubiquity of metaphor, 85-119. Amsterdam: Benjamins.
Dirven, René, and Ralf Pörings. 2002, eds. Metaphor and metonymy in comparison and contrast. Berlin/New York: DeGruyter.

Di Sciullo, Anna-Maria, and Edwin Williams. 1987. On the definition of word. Cambridge, MA: MIT Press.
Dixon, R.M.W. 1980. The languages of Australia. Cambridge: Cambridge University Press.
Dixon, R.M.W. 1982. Problems in Dyirbal dialectology. In: John Anderson (ed.): Language form and linguistic variation. Papers dedicated to Angus McIntosh, 43-73. Amsterdam: Benjamins.
Dixon, R.M.W. 1988. A grammar of Boumaa Fijian. Chicago: University of Chicago Press.
Dixon, R.M.W. 1999. Arawá. In: Dixon and Aikhenvald (eds.), 293-306.
Dixon, R.M.W. 2001. The Australian linguistic area. Areal diffusion and genetic inheritance. In: Aikhenvald and Dixon (eds.), 64-104.
Dixon, R.M.W. 2002. Australian languages: their nature and development. Cambridge: Cambridge University Press.
Dixon, R.M.W. 2010. Basic linguistic theory. Vol. 1: methodology. Oxford: Oxford University Press.
Dixon, R.M.W., and Alexandra Y. Aikhenvald. 1999, eds. The Amazonian languages. Cambridge: Cambridge University Press.
Dixon, R. M.W., and Alan Vogel. 2004. The Jarawara language of Southern Amazonia. Oxford: Oxford University Press.
Djarangar, Djita Issa. 1989. Description phonologique et grammaticale du bedjond, parler Sara de Bediondo Tchad. Lille: ANRT.
Dobrizhoffer, Martin. 1822. An account of the Abipones, an equestrian people of Paraguay. Vol. 2. London: John Murray.
Doherty, Martin J. 2004. Children's difficulty in learning homonyms. Journal of Child Language 31, 203-214.
Donohue, Mark. n.d. a. One dictionary. Manuscript.
Donohue, Mark. n.d. b. Wasur wordlists. Manuscript.
Dorsey, James Owen, and John R. Swanton. 1912. A dictionary of the Biloxi and Ofo languages. Washington: Government Printing Office.
Downing, Pamela. 1977. On the creation and use of English compound nouns. Language 53, 810-842.
Draper, Norm, and Sheila Draper. 2002. Dictionary of Kyaka Enga, Papua New Guinea. Canberra: Pacific Linguistics.
Dryer, Matthew S. 1989. Large linguistic areas and language sampling. Studies in Language 13, 257-92.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. Language 68, 81-138.
Dryer, Matthew S. 2003. Significant and non-significant implicational universals. Linguistic Typology 7, 106-127.
Dryer, Matthew S. 2005a. Genealogical language list. In: Haspelmath et al. (eds.), 584-644.
Dryer, Matthew S. 2005b. Order of adjective and noun. In: Haspelmath et al. (eds.), 354-357.
Dryer, Matthew S. 2005c. Order of demonstrative and noun. In: Haspelmath et al. (eds.), 358-361.
Dryer, Matthew S. 2005d. Order of object and verb. In: Haspelmath et al. (eds.), 338-341.

Dryer, Matthew S. 2005e. Order of relative clause and noun. In: Haspelmath et al. (eds.), 366-369.
Dryer, Matthew S. 2005f. Order of subject and verb. In: Haspelmath et al. (eds.), 334-337.
Dryer, Matthew S. 2005g. Order of subject, object and verb. In: Haspelmath et al. (eds.), 330-333.
Dryer, Matthew S. 2005h. Position of interrogative phrases in content questions. In: Haspelmath et al. (eds.), 378-381.
Dryer, Matthew S. 2005i. Position of polar question particles. In: Haspelmath et al. (eds.), 374-377.
Dryer, Matthew S. 2005j. Relationship between the order of object and verb and the order of adjective and noun. In: Haspelmath et al. (eds.), 394-397.
Dryer, Matthew S. 2005k. Relationship between the order of object and verb and the order of relative clause and noun. In: Haspelmath et al. (eds.), 390-393.
Duanmu, San. 1999. Stress and the development of disyllabic words in Chinese. Diachronica 16, 1-35.
Dunbar, George. 2001. Towards a cognitive analysis of polysemy, ambiguity, and vagueness. Cognitive Linguistics 12, 1-14.
Dunn, Michael John. 1999. A grammar of Chukchi. Canberra: Australian National University.
Durie, Mark, and Malcolm Ross. 1996, eds. The comparative method reviewed: regularity and irregularity in language change. Oxford: Oxford University Press.
Dworkin, Steven N. 1993. La cuasi-homonimia y la perdida lexical en el Español Antiguo. Lexis 17, 57-74.
Dworkin, Steven N. 1993. Near-homonymy, semantic overlap and lexical loss in Medieval Spanish: three case studies. Romanistisches Jahrbuch 44, 271-281.
Eberty, Gustav. 1869. G.B. Vico. Studii critici e comparativi di Carlo Cantoni. (Torino,Civelli 1867). Zeitschrift für Völkerpsychologie und Sprachwissenschaft 6, 429-464.
Eells, Myron. 1889. The Thunder Bird. American Anthropologist 2, 329-336.
Eilers, Wilhelm. 1967. Stern-Planet-Regenbogen. Zur Nomenklatur der Orientalischen Himmelskunde. In: Wilhelm Hoenerbach (ed.): Der Orient in der Forschung. Festschrift für Otto Spies zum 5. April 1966, 92-146. Wiesbaden: Harassowitz.
Eilers, Wilhelm. 1973a. Die Vergleichend-Semasiologische Methode in der Orientalistik. Wiesbaden: Verlag der Wissenschaften und der Literatur in Mainz in Kommission bei Franz Steiner Verlag GmbH.
Eilers, Wilhelm. 1973b. Über Sprache aus der Sicht von Einzelsprachen. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
Einaudi, Paula F. 1974. A grammar of Biloxi. Dissertation, University of Colorado.
Elbert, Samuel. 1953. Internal relationships of Polynesian languages and dialects. Southwestern Journal of Anthropology 9, 147-73.
Elbert, Samuel H. 1979. Hawaiian grammar. Honolulu: University of Hawaii Press.
Elmendorf, William W. 1951. Word taboo and lexical change in Coast Salish. International Journal of American Linguistics 17, 205-208.
Emeneau, M. B. 1948. Taboos on animal names. Language 24, 56-63.

Emeneau, Murray B. 1956. India as a linguistic area. Language 32, 3-16.
Endo, Fubito. 1997. A basic vocabulary of Kolyma Yukaghir. Languages of the North Pacific Rim 2, 147-161.
Endo, Fubito. 2001. Additional basic vocabulary of Kolyma Yukaghir. Languages of the North Pacific Rim 6, 9-27.
Enfield, N.J. 2003. Linguistic epidemiology. Semantics and grammar of language contact in Mainland Southeast Asia. London/New York: RoutledgeCurzon.
Enfield, N.J. 2005. Areal Linguistics and Mainland Southeast Asia. Annual Review of Anthropology 34, 181-206.
Enfield, N. J. 2006. Heterosemy and the grammar-lexicon trade-off. In: Felix Ameka, Alan Dench, and Nicholas Evans (eds.): Catching Language, 297-320. Berlin: De Gruyter.
Enfield, N. J., Asifa Majid, and Miriam van Staden. 2006. Cross-linguistic categorisation of the body: introduction. Language Sciences 28, 137-147.
Enrico, John James. 2005. Haida dictionary. Skidegate, Masset, and Alaskan dialects. Fairbanks: Alaska Native Language Center.
Epps, Patience. 2007. The Vaupés melting pot: Tucanoan influence on Hup. In: Alexandra Y. Aikhenvald and R.M.W. Dixon (eds.): Grammars in contact: a cross-linguistic typology, 267-289. Oxford: Oxford University Press.
Epps, Patience. 2008. A Grammar of Hup. Berlin: De Gruyter.
Epps, Patience. 2009. Hup vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/39
Erdmann, Karl Otto. 1910. Die Bedeutung des Wortes. Aufsätze aus dem Grenzgebiet der Sprachpsychologie und Logik. Second Edition. Leipzig: Avenarius.
Essien, Okon E. 1990. A grammar of the Ibibio language. Ibadan: University Press Ltd.
Estes, Zachary. 2003. Attributive and relational processes in nominal combination. Journal of Memory and Language 48, 304-319.
Estrada Fernández, Zarina. 2009a. Yaqui de Sonora. México, D.F.: Colegio de México.
Estrada Fernández, Zarina. 2009b. Yaqui vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/32
Evans, H. Meurig, and W.O. Thomas. 1983. Y Geiriadur Mawr. The Complete Welsh-English English-Welsh Dictionary. Eleventh Edition. Llandybïe: Christopher Davies (Cyhoeddwyr) Cyf.
Evans, Nicholas. 1990. Without this child – some regularities of semantic change in the Australian linguistic area. In: Peter Austin, R.M.W. Dixon, Tom Dutton, and Isobel White (eds.): Language and history: essays in honour of Luise A. Hercus, 137-155. Canberra: Pacific Linguistics.
Evans, Nicholas. 1992. Multiple semiotic systems, hyperpolysemy, and the reconstruction of semantic change in Australian languages. In: Günter Kellerman and Michael D. Morrissey (eds.): Diachrony within synchrony, 475-508. Bern: Peter Lang.
Evans, Nicholas. 2004. Bininj Gun-Wok. In: Arnaud, Pierre J.L. (ed.): Le nom composé données sur seize langues, 71-99. Lyon: Presses Universitaires de Lyon.
Evans, Nicholas. 2005. Australian languages reconsidered: a review of Dixon (2002). Oceanic Linguistics 44, 242-286.

Evans, Nicholas. 2010. Semantic typology. In: Song (ed.), 504-533.
Evans, Nicholas, and Hans-Jürgen Sasse. 2007. Searching for meaning in the library of Babel: field semantics and problems of digital archiving. Archives and Social Studies: A Journal of Interdisciplinary Research 1, 63-123.
Farr, Cynthia J.M. 2001. Idioms: the darndest things that people say. In: Andrew Pawley, Malcolm Ross, and Darrell Tryon (eds.): The boy from Bundaberg. Studies in Melanesian linguistics in honour of Tom Dutton, 117-141. Canberra: Pacific Linguistics.
Feng, Shengli. 1997. Prosodic structure and compound words in Classical Chinese. In: Packard (ed.), 197-260.
Fernández Garay, Ana. 1998. El Tehuelche: una lengua en vías de extinctión. Valdivia: Estudios Filologicos.
Fernández Garay, Ana. 2004. Diccionario Tehuelche-Español, indice Español-Tehuelche. Leiden: Escuela de Investigación de Estudios Asiáticos.
Feyaerts, Kurt. 2000. Refining the inheritance hypothesis: interaction between metaphoric and metonymic hierarchies. In: Barcelona (ed.), 59-78.
Fillmore, Charles. 1982. Frame semantics. In: Linguistics in the Morning Calm, 111-137. Seoul: Hanshin Publishing.
Firchow, Irwin B. 1970. Form and function of Rotokas words, Bougainville, New Guinea. M.A. thesis. University of California, Davis.
Firchow, Irwin B., and Jaqueline Firchow. 2008. Rotokas-English dictionary. Ukarumpa: Summer Institute of Linguistics.
Fisher, Louise, Wayne Leman, Leroy Pine Sr., and Marie Sanchez. 2008. Cheyenne dictionary. Online, http://cheyenne.110mb.com/.
Foley, William A. 1997. Anthropological linguistics. Oxford: Blackwell.
Fortescue, Michael D. 1980. Affix ordering in West Greenlandic derivational processes. International Journal of American Linguistics 46, 259-278.
Fortescue, Michael D. 1992. Morphophonemic complexity and typological stability in a polysynthetic language family. International Journal of American Lingusitics 58, 242-248.
Fortescue, Michael. 1998. Language relations across Bering Strait. Reappraising the archaeological and linguistic evidence. London / New York: Cassell.
Fortescue, Michael D. 2005. Comparative Chukotko-Kamchatkan dictionary. Berlin: De Gruyter.
Fortescue, Michael, Steven Jacobson, and Lawrence Kaplan. 1994. Comparative Eskimo dictionary. With Aleut cognates. Fairbanks: Alaska Native Language Center.
Fortson, Benjamin W. IV. 2004. Indo-European language and culture. An introduction. Malden: Blackwell.
Fowler, Carol A. 1988. Differential shortening of repeated content words produced in various communicative contexts. Language and Speech 31, 307-319.
François, Alexandre. 2008. Semantic maps and the typology of colexification: intertwining polysemous networks across languages. In: Vanhove (ed.), 163-215.

François, Alexandre. 2010. Des valeurs en héritage: les isomorphismes sémantiques et la reconstruction des langues. In: Injoo Choi-Jonin, Marc Duval, and Olivier Soutet (eds): Typologie et comparatisme, 129-145. Louvain: Peeters.
Fränkel, Ernst. 1938. Beiträge zur Indogermanischen Semasiologie. Revue des Études Indo-Européennes 1, 405-439.
Frantz, Donald G. 1972. The origin of Cheyenne pitch accent. International Journal of American Linguistics 38, 223-25.
Frantz, Donald G. 1991. Blackfoot grammar. Toronto: University of Toronto Press.
Frantz, Donald G., and Norma J. Russell. 1995. Blackfoot dictionary of stems, roots and affixes. Second edition. Toronto: University of Toronto Press.
Frege, Gottlob. 1892. Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik 100, 25-50.
Funke, Otto. 1924/1974. Innere Sprachform: Eine Einführung in A. Martys Sprachphilosophie. Hildesheim: Gerstenberg.
Gabelentz, Georg von der. 1901. Die Sprachwissenschaft. Ihre Aufgaben, Methoden, und bisherigen Ergebnisse. Second Edition. Leipzig: Tauchnitz.
Gagné, Christina L., and Edward J. Shoben. 1997. Influence of thematic relations on the comprehension of modifier-noun combinations. Journal of Experimental Psychology: Learning, Memory, and Cognition 23, 71-87.
Gamkrelidze, Thomas V., and Vjačeslav V. Ivanov. 1995. Indo-European and the Indo-Europeans. A reconstruction and historical analysis of a Proto-Language and Proto-Culture. Berlin: De Gruyter.
Gauger, Hans-Martin. 1971. Durchsichtige Wörter. Zur Theorie der Wortbildung. Heidelberg: Winter.
Geeraerts, Dirk. 1993. Vagueness's puzzles, polysemy's vagaries. Cognitive Linguistics 4, 223-272.
Geeraerts, Dirk. 2002a. The interaction of metaphor and metonymy in composite expressions. In: Dirven and Pörings (eds.), 435-465.
Geeraerts, Dirk. 2002b. The scope of diachronic onomasiology. In: Vilmos Agel, Andreas Gardt, Ulrike Hass-Zumkehr, and Thorsten Roelcke (eds.): Das Wort. Seine strukturelle und kulturelle Dimension. Festschrift für Oskar Reichmann zum 65. Geburtstag, 29-43. Tübingen: Niemeyer.
Geeraerts, Dirk. 2010. Theories of lexical semantics. Oxford: Oxford University Press.
Gentner, Dedre. 1983. Structure-mapping: a theoretical framework for analogy. Cognitive Science 7, 155-170.
Geraghty, Paul. 1986. The sound system of Proto-Central-Pacific. In: Paul Geraghty, Lois Carrington, and S.A. Wurm (eds.): FOCAL II. Papers from the fourth International Conference on Austronesian Linguistics, 289-312. Canberra: Pacific Linguistics.
Gessner, Suzanne. 2003. The prosodic system of the Dakelh (Carrier) language. Dissertation, University of British Columbia.
Gessner, Suzanne, and Gunnar Hansson. 2004. Anti-homophony effects in Dakelh (Carrier) valence morphology. Berkeley Linguistic Society 30, 93-104.

Gibson, James J. 1977. The theory of affordances. In: Robert Shaw and John Bransford (eds.): Perceiving, acting, and knowing: towards an ecological psychology, 67-82. Hillsdale: Lawrence Erlbaum.
Gibson, James J. 1979. The ecological approach to visual perception. Boston: Houghton Mifflin.
Gil, David. 2004. Riau Indonesian *sama*: explorations in macrofunctionality. In: Martin Haspelmath (ed.): Coordinating constructions, 371-424. Amsterdam: Benjamins.
Gil, David. 2005. Para-linguistic usage of clicks. In: Haspelmath et al. (eds.), 572-575.
Gilliéron, J., and M. Roques. 1912. Etudes de géographie linguistique. Paris: Champion.
Givón, Talmy. 1979. On understanding grammar. New York: Academic Press.
Givón, Talmy. 1984. Syntax. A functional-typological introduction. Vol. 1. Amsterdam: Benjamins.
Glasgow, Kathleen. 1994. Burarra - Gun-nartpa dictionary with English finder list. Darwin: Summer Institute of Linguistics.
Glass, Amee, and Dorothy Hackett. 2003. Ngaanyatjarra-Ngaatjatjarra to English dictionary. Alice Springs: IAD Press.
Goddard, Cliff. 2001. Lexico-semantic universals: a critical overview. Linguistic Typology 5, 1-65.
Goddard, Cliff. 1994. Semantic theory and semantic universals. In: Goddard and Wierzbicka (eds.), 7-30.
Goddard, Cliff, and Anna Wierzbicka 1994, eds. Semantic and lexical universals: theory and empirical findings. Amsterdam: Benjamins.
Goddard, Cliff, and Anna Wierzbicka. 2010. Semantics and cognition. Wiley Interdisciplinary Reviews: Cognitive Science 2, 125-135.
Goddard, Ives. 1990. Algonquian linguistic change and reconstruction. In: Philip Baldi (ed.): Linguistic change and reconstruction methodology, 99-114. Berlin: De Gruyter.
Goldie, Hugh. 1964. Dictionary of the Efik language. Farnborough: Gregg Press.
Gómez Bacarreza, Donato, and José Condori Cosme. 1992. Morfología y gramatica del idioma Aymara. La Paz.
Gómez Rendón, Jorge A. 2009. Imbabura Quechua vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/37.
Gonçalves, Cristina Helena R. C. 1987. Concordância em Mundurukú. Campinas: Editora de Unicamp.
Goodwin Gómez, Gale. 2000. Noun classifiers in ethnobotanical terminology of a Yanomami language of Brazil. In: Hein van der Voort and Simon van de Kerke (eds.): Indigenous Languages of Lowland South America, 9-24. Leiden: CNWS.
Goossens, Louis. 1990. Metaphtonymy: the interaction of metaphor and metonymy in expressions for linguistic action. Cognitive Linguistics 1, 323-340.
Grady, Joseph E. 1999. A typology of motivation for conceptual metaphor: correlation vs. resemblance. In: Raymond W. Gibbs and Gerard J. Steen (eds.): Metaphor in Cognitive Linguistics, 79-100. Amsterdam: Benjamins.

Gravelle, Gilles. 1998. Syntactic constructions and the Meyah Lexicon. In: Jelle Miedema, Cecilia Odé, Rien A.C. Dam, and Connie Baak (eds.): Perspectives on the Bird's Head of Irian Jaya, Indonesia: Proceedings of the Conference, Leiden, 13-17 October 1997, 555-573. Amsterdam: Editions Rodopi B.V.

Gravelle, Gilles. 2004. Meyah: an East Bird's Head language of Papua, Indonesia. Ph.D. dissertation, Vrije Universiteit Amsterdam.

Green, Rebecca. 1987. A sketch grammar of Burarra. B.A. Honours Thesis, Australian National University.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. International Journal of American Linguistics 26, 178-194.

Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In: Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.): Universals of human language I: method and theory, 61-92. Stanford: Stanford University Press.

Greenberg, Joseph H. 1995. The diachronic typological approach to language. In: Masayoshi Shibatani and Theodora Bynon (eds.): Approaches to language typology, 143-166. Oxford: Clarendon.

Grenand, Françoise. 1980. La langue Wayãpi (Guyane Française). Phonologie et grammaire. Paris: Centre National de la Recherche Scientifique.

Grenand, Françoise. 1989. Diccionaire Wayãpi-Française. Paris: Peeters/SELAF.

Grosh, Sylvia, and Andrew Grosh. 2005. Organised phonology data supplement: Kaluli language. In: Steve Parker (ed.): Phonological descriptions of PNG languages, 57-74. Ukarumpa: Summer Institute of Linguistics.

Gruzdeva, Ekaterina. 1998. Nivkh. Munich: Lincom Europa.

Grzega, Joachim. 2009. Bibliography of onomasiological works. Online: http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOn-7.pdf.

Guasch, Antonio, and Diego Ortiz. 1998. Diccionario Castellano-Guaraní, Guaraní-Castellano. Asuncion: Centro de Estudios Paraguayos 'Antonio Guasch'.

Guillaume, Antoine. 2008. A grammar of Cavineña. Berlin: De Gruyter.

Gumperz, John J., and Robert Wilson. 1971. Convergence and creolization: a case study from the Indo-Aryan/Dravidian border in India. In: Hymes (ed.), 151-167.

Haacke, Wilfried H.G., and Eliphas Eiseb. 2002. A Khoekhoegowab Dictionary with an English-Khoekhoegowab Index. Windhoek: Gamsberg Macmillan.

Hagege, Claude. 1970. La langue Mbum de Nganha (Cameroun): phonologie, grammaire. Paris: SELAF.

Hagège, Claude. 1993. The language builder. An essay on the human signature in linguistic morphogenesis. Amsterdam: Benjamins.

Hagman, Roy S. 1977. Nama Hottentot grammar. Bloomington: Indiana University.

Haiman, John. 1974. Concessives, conditionals, and verbs of volition. Foundations of Language 11, 341-359.

Haiman, John. 1985. Natural syntax. Iconicity and erosion. Cambridge: Cambridge University Press.

Hajek, John. 2004. Consonant inventories as an areal feature of the New Guinea-Pacific region: testing Trudgill's hypothesis. Linguistic Typology 8, 343-350.

Hale, Kenneth. 1994. Preliminary observations on lexical and semantic primitives in the Misumalpan languages of Nicaragua. In: Goddard and Wierzbicka (eds.), 263-283.
Hall, Robert L. 1997. An archaeology of the soul: North American Indian belief and ritual. Champaign: University of Illinois Press.
Hammarström, Harald. 2010. A full-scale test of the language farming dispersal hypothesis. Diachronica 27, 197-213.
Hanks, Patrick. 2000. Do word meanings exist? Computers and the Humanities 34, 205-215.
Harcourt, Raoul d'. 1962/2002. Textiles of Ancient Peru and their techniques. Ed. by Grace G. Denny and Carolyn M. Osborne, transl. by Sadie Brown. Mincola: Dover Press.
Hardman, Martha J. 2001. Aymara. Munich: Lincom Europa.
Harrington, John P. 1928. Vocabulary of the Kiowa Language. Washington: United States Government Printing Office.
Harris, Roy. 1983. Course in general linguistics. London: G. Duckworth. [Translation of Saussure 1916/1967]
Harrison, Roy, Margaret Harrison, and Cástulo H. Garcia. 1981. Diccionario Zoque de Copainalá. México, D.F.: Instituto Lingüístico de Verano.
Hart Elton, George, Helen Long de Hart, and Esther Gordon de Powlinson. 1976. La fonologia del Chayahuita. Lima: Instituto Lingüístico de Verano.
Hart, Helen. 2008. Diccionario Chayahuita-Castellano. Secunda Edición. Pucallpa: Instituto Lingüístico de Verano.
Haspelmath, Martin. 1999. Optimality and diachronic adaptation. Zeitschrift für Sprachwissenschaft 18, 180-205.
Haspelmath, Martin. 2002. Understanding morphology. London: Arnold.
Haspelmath, Martin. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In: Michael Tomasello (ed.): The new psychology of language, Vol. 2, 211-242. Mahwah, NJ: Lawrence Erlbaum.
Haspelmath, Martin. 2005. Occurrence of nominal plurality. In: Haspelmath et al. (eds.), 142-145.
Haspelmath, Martin. 2007. Pre-established categories don't exist: consequences for language description and typology. Linguistic Typology 11, 119-132.
Haspelmath, Martin. 2009. Lexical borrowings: concepts and issues. In: Haspelmath and Tadmor (eds.), 35-54.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. Language 86, 663-687.
Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible. 2001, eds. Language typology and language universals. Berlin: De Gruyter.
Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie. 2005, eds. The world atlas of language structures. Oxford: Oxford University Press.
Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie. 2008, eds. The world atlas of language structures online. Munich: Max Planck Digital Library.
Haspelmath, Martin, and Uri Tadmor. 2009a, eds. Loanwords in the world's languages: a comparative handbook. Berlin: De Gruyter.

Haspelmath, Martin, and Uri Tadmor. 2009b. The loanword typology project and the world loanword database. In: Haspelmath and Tadmor 2009a (eds)., 1-34.
Haspelmath, Martin, and Uri Tadmor. 2009c, eds. The world loanword database. www.wold.livingsources.org. Munich: Max Planck Digital Library.
Haudricourt, André G. 1954. Comment reconstruire le chinois archaique. Word 10, 351-364.
Haudricourt, André G. 1956. De la restitution des initials dans les langues monosyllabiques: le problème du Thai Commun. Bulletin de la Société de Linguistique de Paris 52, 307-322.
Haudricourt, André G. 1968. Linguistique et ethnologie. In : Jean Poirier (ed.): Ethnologie générale. Encyclopédie de la Pléiade, 288-316. Paris: Gallimard.
Haugen, Einar. 1950. The analysis of linguistic borrowing. Language 26, 210-231.
Haugen, Einar. 1956. Bilingualism in the Americas. A bibliography and research guide. Tuscaloosa: University of Alabama Press.
Hayes, Patrick J. 1985. The second naive physics manifesto. In: Jerry R. Hobbs and Robert C. Moore (eds.): Formal theories of the commonsense world, 1-36. Norwood: Ablex.
Hayne, Thomas. 1639. Linguarum cognatio. Seu de linguis in genere, & de variarum linguarum harmoniā dissertatio. London: John Norton.
Heath, G.R. 1950. Miskito glossary, with ethnographic commentary. International Journal of American Linguistics 16, 20-34.
Heath, George R., and Werner G. Marx. 1961. Diccionario Miskito-Español, Español-Miskito. Tegucigalpa: Imoprenta Calderon.
Heath, Jeffrey. 1981. A case of extensive lexical diffusion: Arnhem Land, Australia. Language 57, 335-367.
Heath, Jeffrey. 1982. Nunggubuyu dictionary. Canberra: Australian Institute of Aboriginal Studies.
Heath, Jeffrey. 1984. Functional grammar of Nunggubuyu. Canberra: Australian Institute of Aboriginal Studies.
Heath, Jeffrey. 1998. Dictionnaire Songhay-Anglais-Français. Tome III: Koroboro Senni. Paris: L'Harmattan.
Heath, Jeffrey. 1999. A grammar of Koyraboro (Koroboro) Senni. The Songhay of Gao, Mali. Cologne: Köppe.
Heath, Jeffrey, and Laura McPherson. 2009. Cognitive set and lexicalization strategy in Dogon action verbs. Anthropological Linguistics 51, 38-63.
Heeschen, Volker. 1992. A Dictionary of the Yale (Kosarek) language (with sketch of grammar and English index). Berlin: Reimer.
Heine, Bernd. 1982. The Nubi language of Kibera – an Arabic creole. Berlin: Reimer.
Heine, Bernd. 1997. Cognitive foundations of grammar. Oxford: Oxford University Press.
Heine, Bernd, and Tania Kuteva. 2002. World lexicon of grammaticalization. Cambridge: Cambridge University Press.
Heine, Bernd, and Tania Kuteva. 2005. Language contact and grammatical change. Cambridge: Cambridge University Press.
Helmbrecht, Johannes. 2005. Politeness distinctions in pronouns. In: Haspelmath et al. (eds.), 186-189.

Hénault-Sakhno, Christine, and Serguei Sakhno. 2005. Typologie sémantique lexicale: problèmes de systématisation. In: Gilbert Lazard and Claire Moyse-Faurie (eds.): Linguistique typologique, 71-90. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Henderson, Eugenie J.A. 1997. Bwe Karen dictionary. With texts and English-Karen word list. London: School of Oriental and African Studies, University of London.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? Behavioural and Brain Sciences 33, 61-83.

Herbermann, Clemens-Peter. 2002. Das Wort als lexikalische Einheit. In: Cruse et al. (eds), 14-33.

Hewitt, George. 2005. North West Caucasian. Lingua 115, 91-145.

Hewitt, George. 2008. Are verbs always what they seem to be? Iran and the Caucasus 12, 307-323.

Hildebrandt, Kristine A. 2003. Manange tone: scenarios of retention and loss in two communities. Dissertation, University of California at Santa Barbara.

Hildebrandt, Kristine A. 2009. Manange vocabulary. In: Haspelmath and Tadmor (eds.). http://wold.livingsources.org/vocabulary/17.

Hilpert, Martin. 2007. Chained metonymies in lexicon and grammar. A cross-linguistic perspective on body part terms. In: Günter Radden, Klaus-Michael Köpcke, Thomas Berg, and Peter Siemund (eds.): Aspects of meaning construction, 77-98. Amsterdam: Benjamins.

Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticization: opposite or orthogonal? In: Walter Bisang, Nikolaus P. Himmelmann, and Björn Wiemer (eds): What makes grammaticalization? A look from its components and its fringes, 21-42. Berlin: De Gruyter.

Hino, Shun'ya. 1978. The classified vocabulary of the Mbum language in Mbang Mboum. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).

Hjelmslev, Louis. 1963. Prolegomena to a theory of language. Madison: University of Wisconsin Press.

Hladký, Josef. 1986. The Czech and the English names of mushrooms. Brno: Masarykova Univerzita v Brně.

Hockett, Charles F. 1958. A course in modern linguistics. New York: Macmillan.

Hockings, Paul, and Christiane Pilot-Raichoor. 1992. A Badaga-English dictionary. Berlin: De Gruyter.

Hoeksema, Jakob. 2000. Compositionality of meaning. In: Geert E. Booij, Christian Lehmann, and Joachim Mugdan (eds.): Morphology, 851-857. Berlin: De Gruyter.

Hofling, Charles Andrew, and Félix Fernando Tesucún. 1997. Itzaj Maya-Spanish-English dictionary. Salt Lake City: University of Utah Press.

Hofling, Charles Andrew, and Félix Fernando Tesucún. 2000. Itzaj Maya grammar. Salt Lake City: University of Utah Press.

Holm, John A. 1988. Pidgin and Creole languages. Cambridge: Cambridge University Press.

Holm, John A. 2000. An introduction to Pidgins and Creoles. Cambridge: Cambridge University Press.

Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List, and Dmitry Egorov. 2011. Automated dating of the world's language families based on lexical similarity. Current Anthropology 52, 841-875.

Holmer, Nils M. 1966. Oceanic semantics (a study in the framing of concepts in the native languages of Australia and Oceania). Upsala: A.-B. Lundequistska Bokhandeln.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. Phonetic explanations for the development of tones. Language 55, 37-58.

Hopper, Paul J., and Elizabeth Closs Traugott. 2002. Grammaticalization. Second Edition. Cambridge: Cambridge University Press.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: Deborah Schiffrin (ed.): Meaning, form, and use in context: linguistic applications, 11-42. Washington: Georgetown University Press.

Hudak, Thomas John. 1991. William J. Gedney's the Yay language: glossary, texts, and translations. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Hull, Geoffrey. 2001. Standard Tetum-English dictionary. Second Edition. Sydney: Allen & Unwin.

Hull, Geoffrey, and Lance Eccles. 2004. Tetum Reference Grammar. Revised Edition. Winston Hills: Sebastião Aparício da Silva Project.

Humboldt, Wilhelm von. 1963/1821. Essai sur les langues du nouveau continent. In: Andreas Flitner and Klaus Giel (eds.): Wilhelm von Humboldt. Werke in fünf Bänden. Vol. 3.: Schriften zur Sprachphilosophie, 300-341. Darmstadt: Wissenschaftliche Buchgesellschaft.

Humboldt, Wilhelm von. 1998/1832. Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts. München: UTB.

Hutchison, John P. 1981. A reference grammar of the Kanuri Language. Madison: University of Wisconsin-Madison.

Huttar, George L. 1975. Sources of creole semantic structures. Language 51, 684-695.

Huttar, George L. 2003. Scales of basicness in semantic domains and their application to creolization. In: Mary Ruth Wise, Thomas N. Headland and Ruth M. Brend (eds.): Language and life: essays in memory of Kenneth L. Pike, 119-137. Dallas: SIL International and the University of Texas at Arlington.

Hyman, Larry M. 1981. Noni grammatical structure. With special reference to verb morphology. Los Angeles: Department of Linguistics, University of Southern C alifornia.

Hymes, Dell. 1971, ed. Pidginization and creolization of languages. Cambridge: Cambridge University Press.

Immler, Manfred. 1991. Is semantics universal, or isn't it? On the relation of language, thought and semantic structure. In: Zaefferer (ed.), 37-59.
Ingham, Bruce. 2003. Lakota. Munich: Lincom Europa.
Inhoff, Albrecht W., Deborah Briihl, and Jill Schwartz. 1996. Compound word effects differ in reading, on-line naming, and delayed naming tasks. Memory and Cognition 24, 466-476.
Inia, Elizabeth K., Sofie Arntsen, Hans Schmidt, Jan Rensel, and Alan Howard 1998. A new Rotuman dictionary. An English Rotuman wordlist. Suva: University of the South Pacific.
Irwin, Geoffrey J. 1992. The prehistoric exploration and colonisation of the Pacific. Cambridge: Cambridge University Press.
Iverson, Gregory K., and Joseph C. Salmons. 1992. The phonology of the Proto-Indo-European root structure constraints. Lingua 87, 293-320.
Jackendoff, Ray. 2002. What's in the Lexicon? In: Sieb Nooteboom, Fred Weerman, and Frank Wijnen (eds.): Storage and computation in the language faculty, 23-58. Dordrecht: Kluwer.
Jacobson, Steven A. 1984. Yup'ik Eskimo dictionary. Fairbanks: Alaska Native American Center.
Jakway, Martha, Horacio C. Lorenzo, and Antonia A. Arturo. 1987. Vocabulario Huambisa. Yarinacocha: Instituto Lingüístico de Verano.
Janhunen, Juha. 2001. Indo-Uralic and Ural-Altaic: on the diachronic implications of areal typology. In: Christian Carpelan, Asko Parpola, and Petteri Koskikallio (eds.): Early contacts between Uralic and Indo-European. Linguistic and archaeological considerations, 207-220. Helsinki: Suomalais-Ugrilainen Seura.
Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de la Société Finno-Ougrienne 258, 57-78.
Jelinek, Eloise. 1984. Empty categories, case, and configurationality. Natural Language and Linguistic Theory 2, 39-76.
Jensen, Cheryl. 1999. Tupí-Guaraní. In: Dixon and Aikhenvald (eds.), 125-163.
Jespersen, Otto. 1942. A Modern English grammar on historical principles. Vol. 6: morphology. Copenhagen: Munksgaard.
Ji, Li-Jun, Zhiyong Zhang, and Richard E. Nisbett. 2004. Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. Journal of Personality and Social Psychology 87, 57-65.
Jóhanneson, Alexander. 1949. Origin of language. Four essays. Reykjavík: H.F. Leiftur.
Johnson, Allen W., and Timothy Earle. 2000. The evolution of human societies: from foraging group to agrarian state. Second edition. Stanford: Stanford University Press.
Jorgensen, Julia C. 1990. The psychological reality of word senses. Journal of Psycholinguistic Research 19, 167-190.
Joseph, Brian D., and Irene Philippaki-Warburton. 1987. Modern Greek. London: Routledge.
Joseph, Brian D., and Richard D. Janda. 2003, eds. The handbook of historical linguistics. Malden: Blackwell.

Jucquois, Guy. 1966. La structure des racines en indo-européen envisagée d'un point de vue statistique. In: Yvan Lebrun (ed.): Linguistic research in Belgium, 57-68. Wetteren: Universa.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. Language 72, 533-578.

Justeson, John S., and Laurence D. Stephens. 1984. On the relationship between the numbers of vowels and consonants in phonological systems. Linguistics 22, 531-546.

Kagaya, Ryohei. 1992. A classified vocabulary of the Bakueri language. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (Ilcaa).

Kastovsky, Dieter. 1982. Wortbildung und Semantik. Düsseldorf: Schwann-Bagel/ Bern: Francke.

Kastovsky, Dieter. 2006a. Historical morphology from a typological point of view. Examples from English. In: Terttu Nevalainen, Juhani Klemola, and Mikko Laitinen (eds.): Types of variation. Diachronic, dialectal and typological Interfaces, 53-80. Amsterdam: Benjamins.

Kastovsky, Dieter. 2006b. Typological changes in derivational morphology. In: Ans van Kemenade and Bettelou Los (eds.): The handbook of the history of English, 151-176. Malden: Blackwell.

Kelso, J.A. Scott. 1995. Dynamic patterns. The self-organization of brain and behaviour. Cambridge, Ma.: MIT Press.

Kemler Nelson, Deborah G., Lindsay Herron, and Morghan B. Holt. 2003. The sources of young children's name innovations for novel artifacts. Journal of Child Language 30, 823-843.

Kemo, Gideon, Loisa Mallo Hanasbey, Abraham Rollo, Alfius Mallo, and Mark Donohue. 2002. Rópu we Te Máwo pílang te. Skou dictionary draft. Manuscript.

Kibrik, Andrej A. 1986. The meaning-form correspondence in grammatical description. In: Winfred P. Lehmann (ed.): Language Typology 1985, 163-168. Amsterdam: Benjamins.

Kibrik, Andrej A. 1992. Relativization in polysynthetic languages. International Journal of American Linguistics 58, 135-157.

Kieft, Anthonij P. 1938. Homonymie en haar invloed op de taalontwikkeling. Groningen: Wolters.

Kilgarriff, Adam. 1997. 'I don't believe in word senses'. Computers and the Humanities 31, 91–113.

King, Robert D. 1967. Functional load and sound change. Language 43, 831-852.

Kinkade, Dale M. 1963. Phonology and morphology of Upper Chehalis. Dissertation, Indiana University.

Kinkade, M. Dale. 1991. Upper Chehalis dictionary. Missoula: University of Montana.

Klimov, G. A. 1977. Tipologiya yazykov aktivnogo stroya [Typology of languages of active structure]. Moscow: Nauka.

Kluge, Friedrich. 2002. Etymologisches Wörterbuch der Deutschen Sprache. 24th edition, ed.by Elmar Seebold. Berlin: De Gruyter.

Koch, Peter. 2001. Lexical typology from a cognitive and linguistic point of view. In: Haspelmath et al. (eds.), 1142-1178.
Koch, Peter. 2008. Cognitive onomasiology and lexical change: around the eye. In: Vanhove (ed.), 107-137.
Koch, Peter, and Daniela Marzo. 2007. A two-dimensional approach to the study of motivation in lexical typology and its first application to French high-frequency vocabulary. Studies in Language 31, 259-291.
Koerner, Ernst Frideryk Konrad. 1971. Ferdinand de Saussure: origin and development of his linguistic theory in western studies of languages. Dissertation, Simon Fraser University.
Koptjevskaja-Tamm, Maria. 2008. Approaching lexical typology. In: Vanhove (ed.), 3-52.
Kortmann, Bernd, and Benedikt Szmrecsanyi. 2009. World Englishes between simplification and complexification. In: Lucia Siebers and Thomas Hoffmann (eds): World Englishes. Problems, properties and prospects, 265-285. Amsterdam: Benjamins.
Kövecses, Zoltán. 2002. Metaphor. A practical introduction. Oxford: Oxford University Press.
Köveces, Zoltán. 2006. Language, mind, and culture. A practical introduction. Oxford: Oxford University Press.
Köveces, Zoltán, and Günter Radden. 1998. Metonymy: developing a Cognitive Linguistics view. Cognitive Linguistics 9, 37-78.
Kramer, Johannes. 2000. Die frühe Entwicklung des onomasiologischen Ansatzes in der Sprachwissenschaft und Lexikographie des 19. Jahrhunderts, In: Auroux et al. (eds.), 1611-1617.
Kreger, Glenn Albert, and Emily Florence Scharfe de Stairs. 1981. Diccionario Huave de San Mateo del Mar. México, D.F.: Instituto Lingüístico de Verano.
Kroeber, Alfred L. 1925/1976. Handbook of the Indians of California. New York: Dover.
Kröger, Franz. 1992. Buli-English dictionary. Münster/Hamburg: Lit Verlag.
Kroneman, Margreet, and Rima Duha. 2005. Sam ey go Nebut. Kamus Bergambar. Manuscript, Summer Institute of Linguistics.
Krupa, Viktor. 1966. Morpheme and word in Maori. The Hague: Mouton.
Kruszewski, Mikolai. 1890. Prinzipien der Sprachentwickelung. Internationale Zeitschrift für Allgemeine Sprachwissenschaft 5, 133-144.
Kuipers, Aert H. 1960. Phoneme and morpheme in Kabardian (Eastern Adyghe). 'S-Gravenhage: Mouton & Co.
Kurebito, Tokusu. 2001. Comparative basic vocabulary of the Chukchee-Kamchatkan language family. Sravnitel'nyj bazovyj slovar' jazykov cukotsko-kamcatskoj jazykovoj sem'. Osaka: ELPR.
Lakoff, George. 1987. Women, fire, and dangerous things: what categories reveal about the mind. Chicago: University of Chicago Press.
Lakoff, George, and Mark Johnson. 1980. Metaphors we live by. Chicago: University of Chicago Press.

Langacker, Ronald W. 1987a. Foundations of cognitive grammar: theoretical prerequisites. Stanford: Stanford University Press.
Langacker, Ronald. 1987b. Nouns and verbs. Language 63, 53-94.
Lass, Roger. 1994. On 'root-based' Indo-European: an embryological or phylogenetic note. Views 3, 31-34.
Laycock, Donald C. 1970. Eliciting basic vocabulary in New Guinea. In: S.A. Wurm and D. C. Laycock (eds.): Pacific linguistic studies in honour of Arthur Capell, 1127-1176. Canberra: Pacific Linguistics.
Laycock, Donald C. 1975. Observations on number systems and semantics. In: S. A. Wurm (ed.): New Guinea area languages and language study. Vol 1: Papuan languages and the New Guinea linguistic scene, 219-233. Canberra: Australian National University.
Laycock, Donald C. 2003. A dictionary of Buin, a language of Bougainville. Ed. by Masayuki Onishi. Canberra: Australian National University.
Lazarus, Moritz, and Heymann Steinthal. 1860. Einleitende Gedanken über Völkerpſychologie, als Einladung zu einer Zeitſchrift für Völkerpſychologie und Sprachwiſſenschaft. Zeitschrift für Völkerpſychologie und Sprachwiſſenschaft 1, 1-73.
Lee, Felicia. 1999. Antisymmetry and the syntax of San Lucas Quiaviní Zapotec. Dissertation, University of California at Los Angeles.
Lefebvre, Claire. 1998. Creole genesis and the acquisition of grammar. The case of Haitian Creole. Cambridge: Cambridge University Press.
Lefebvre, Claire. 2001. Relexification. A process available to human cognition. Berkeley Linguistics Society 27, 125-139.
Lehmann, Christian. 1990. Towards lexical typology. In: William Croft, Suzanne Kemmer, and Keith Denning (eds.): Studies in typology and diachrony. Papers presented to Joseph H. Greenberg on his 75th birthday, 161-185. Amsterdam/ Philadelphia: Benjamins.
Lehmann, Winfred P. 2002. Pre-Indo-European. Washington: Institute for the Study of Man.
Lehrer, Adrienne, and Eva Kittay 1992, eds. Frames, fields, and contrasts: new essays in semantic and lexical organization. Hillsdale: Lawrence Erlbaum.
Lessing, Ferdinand D. 1995, ed. Mongolian-English-dictionary. 3rd printing. Bloomington: Indiana University.
Levi, Judith N. 1978. The syntax and semantics of complex nominals. New York: Academic Press.
Levine, Robert D. 1977. The Skidegate dialect of Haida. Dissertation, University of New York.
Levinson, David. 1991, ed. The encyclopedia of world cultures. Boston: Hall.
Levinson, Stephen C. 1994. Vision, shape, and linguistic description: Tzeltal body-part terminology and object description. Linguistics 32, 791-855.

Levinson, Stephen C. 2006a. Introduction: the evolution of culture in a microcosm. In: Stephen C. Levinson and Pierre Jaisson (eds.): Evolution and culture: a Fyssen Foundation symposium, 1-41. Cambridge: MIT Press.

Levinson, Stephen C. 2006b. Parts of the body in Yélî Dnye, the Papuan language of Rossel Island. Language Sciences 28, 221-240.

Levinson, Stephen C. 2008. Landscape, seascape and the ontology of places on Rossel Island, Papua New Guinea. Language Sciences 30, 256-290.

Levinson, Stephen C., and Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. Language 79, 485-516.

Lewis, M. Paul. 2009, ed. Ethnologue: languages of the world. 16th edition. Dallas: SIL International.

Lewis, Paul W., and Bai Bibo. 1997. Hani-English English-Hani dictionary: aqniqdoq-Yilydoq doqlo-soqdaoq. London: Paul Kegan.

Li, Charles N., and Sandra A.Thompson. 1981. Mandarin Chinese. A functional reference grammar. Berkeley/Los Angeles: University of California Press.

Libben, Gary. 2003. Morphological parsing and morphological structure. In: Egbert M.H. Assink and Dominiek Sandra (eds.): Reading complex words. Cross-language studies, 221-239. New York: Kluwer/Plenum.

Libben, Gary. 2006. Why study compound processing? An overview of the issues. In: Libben and Jarema (eds)., 1-23.

Libben, Gary, and Gonia Jarema. 2006, eds. The representation and processing of compound words. Oxford: Oxford University Press.

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. Language 67, 475-509.

Lieber, Michael D., and Kalio H. Dikepa. 1974. Kapingamarangi lexicon. Honululu: University of Hawaii Press.

Lieber, Rochelle, and Pavol Štekauer. 2009, eds. The Oxford handbook of compounding. Oxford: Oxford University Press.

Lin, Hua. 2001. A grammar of Mandarin Chinese. Munich: Lincom Europa.

Lindblom, Björn. 1998. Systemic constraints and adaptive change in the formation of sound structure. In: James A. Hurford, Michael Studdert-Kennedy, and Chris Knight (eds): Approaches to the evolution of language, 242-264. Cambridge: Cambridge University Press.

Lindblom, Björn. 2000. Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns. Phonetica 57, 297-314.

Lindskoog, John N., and Carrie A. Lindskoog. 1964. Vocabulario Cayapa. Quito: Instituto Lingüístico de Verano.

Litteral, Robert Lee. 1980. Features of Anggor discourse. Dissertation, University of Pennsylvania.

Litteral, Shirley (deceased). n.d. Anggor dictionary. Manuscript, Summer Institute of Linguistics.

Lizot, Jacques. 1975. Diccionario Yanomami-Español. Cáracas: Universidad Central de Venezuela.
Lloyd, J., and A. Healey. 1970. Barua phonemes: a problem in interpretation. Linguistics 60, 33-48.
Lloyd, John A. 1992. A Baruya-Tok Pisin-English dictionary. Canberra: Australian National University.
Löhr, Doris, H. Ekkehard Wolff, and Ari Awagana. 2009. Kanuri vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/5.
Losonsky, Michael. 1999. On language. On the diversity of human language construction and its influence on the mental development of the human species. Cambridge: Cambridge University Press. [Translation of Humboldt 1832/1998]
Lowes, R.H.G. 1954. Alphabetical list of Lengua Indian words with English equivalents (Paraguayan Chaco). Journal de la Société des Américanistes de Paris 43, 86-107.
Lucy, John A. 1992. Grammatical categories and cognition: a case study of the Linguistic Relativity Hypothesis. Cambridge: Cambridge University Press.
Lupyan Gary, and Rick Dale. 2010. Language structure is partly determined by social structure. PLoS ONE 5: e8559.
Lux, David. 2003. Noni provisional lexicon. Manuscript, Summer Institute of Linguistics.
Lynch, John. 1977. Lenakel dictionary. Canberra: Australian National University.
Lynch, John. 1978. A grammar of Lenakel. Canberra: Australian National University.
Lynch, John, Malcolm Ross, and Terry Crowley. 2002a. Proto-Oceanic. In: Lynch et al. (eds.), 54-91.
Lynch, John, Malcolm Ross, and Terry Crowley. 2002b, eds. The Oceanic languages. London: Routledge.
Lyons, John. 1966. Towards a 'notional' theory of the 'parts of speech'. Journal of Linguistics 2, 209-236.
Lyons, John. 1977. Semantics. 2 vols. Cambridge: Cambridge University Press.
Lyovin, Anatole. 1977. Sound change, homophony, and lexical diffusion. In: William S.-Y. Wang (ed.): The lexicon in phonological change, 120- 132. The Hague: Mouton.
Maddieson, Ian. 1984. Patterns of sound. Cambridge: Cambridge University Press.
Maddieson, Ian. 2005a. Consonant inventories. In: Haspelmath et al. (eds.), 10-13.
Maddieson, Ian. 2005b. Consonant-vowel ratio. In: Haspelmath et al. (eds.), 18-21.
Maddieson, Ian. 2005c. Lateral consonants. In: Haspelmath et al. (eds.), 38-41.
Maddieson, Ian. 2005d. Syllable structure. In: Haspelmath et al. (eds.), 54-57.
Maddieson, Ian. 2005e. Tone. In: Haspelmath et al. (eds.), 58-61.
Maddieson, Ian. 2005f. Uvular consonants. In: Haspelmath et al. (eds.), 30-33.
Maddieson, Ian. 2005g. Voicing in plosives and fricatives. In: Haspelmath et al. (eds.), 22-25.
Maddieson, Ian. 2005h. Vowel quality inventories. In: Haspelmath et al. (eds.), 14-17.
Malinowski, Bronislaw. 1935. Coral gardens and their magic: a study of the methods of tilling the soil and of agricultural rites in the Trobriand Islands. London: Routledge.

Malkiel, Yakov. 1952. Studies in Hispano-Latin homonymics: pessulus, pactus, pectus, dēspectus, suspectus, fistula in Ibero-Romance. Language 28, 299-338.
Malkiel, Yakov. 1979. Problems in the diachronic differentiation of near-homophones. Language 55, 1-36.
Mallory, J. P., and D. Q. Adams. 2006. The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world. Oxford: Oxford University Press.
Man, Edward H. 1923. Dictionary of the South Andaman language. Bombay: British India Press.
Manoharan, S. 1986. Linguistic peculiarities of Andamanese family of languages. Indian Linguistics 47, 25-32.
Manoharan, S. 1989. A descriptive and comparative study of Andamanese language. Calcutta: Anthropological Survey of India, Government of India, Ministry of Human Resource Development.
Marchand, Hans. 1969. The categories and types of present-day English word formation. A synchronic-diachronic approach. Munich: C.H. Beck.
Maring, Joel M. 1967. A grammar of Acoma Keresan. Dissertation, Indiana University.
Mark, David M., Barry Smith, and Barbara Tversky. 1999. Ontology and geographic objects: an empirical study of cognitive categorization. In: Christian Freksa and David M. Mark (eds.): Spatial information theory: cognitive and computational foundations of geographic information science, 283-298. Berlin/Heidelberg/New York: Springer.
Mark, David M., and Andrew G. Turk. 2003a. Ethnophysiography. Manuscript.
Mark, David M., and Andrew G. Turk. 2003b. Landscape categories in Yindjibarndi: ontology, environment and language. In: Werner Kuhn, Michael Worboys, and Sabine Timpf (eds.): Spatial Information Theory. Foundation of geographic information science, 28-45. Berlin/Heidelberg/New York: Springer.
Markel, Stephen. 1995. Origins of the Indian planetary deities. Lewiston: Edwin Mellen Press.
Martinet, André. 1952. Function, structure, and sound change. Word 8, 1-32.
Marty, Anton. 1908. Untersuchungen zur Grundlegung der allgemeinen Grammatik und Sprachphilosophie. Halle: Niemeyer.
Marty, Anton. 1950. Über Wert und Methode einer allgemeinen beschreibenden Bedeutungslehre. Ed. by Otto Funke. Bern: Francke.
Marzo, Daniela, and Verena Rube. 2006. What do you think where words come from? Investigating lexical motivation empirically. In: Valery Solovyev, Vera Goldberg, and Vladimir Polyakov (eds.): The VIII-th International Conference Cognitive Modeling in Linguistics. Proceedings. Vol.I, 152-161. Moscow: ZAO.
Marzo, Daniela, Verena Rube, and Birgit Umbreit. 2006. Lexical motivation and speaker judgements. Poster, Linguistic Evidence II Conference, Tübingen, 2.-4.2.2006.

Masica, Colin P. 2001. The definition and significance of linguistic areas. Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In: Peri Bhaskararao and Karumuri V. Subbarao (eds.): The yearbook of South Asian languages and linguistics 2001. Papers from the symposium on South Asian languages: contact, convergence and typology, 205-267. New Delhi: Sage Publications.

Maslova, Elena. 1998. A grammar of Kolyma Yukaghir. Dissertation, University of Bielefeld.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. Linguistic Typology 4, 307-333.

Masuda, Takahiko, and Richard E. Nisbett. 2001. Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. Journal of Personality and Social Psychology 81, 922-934.

Mather, Elsie, Marie Meade, and Osahito Miyaoka. 2002. Survey of Yup'ik grammar. Kyoto: ELPR.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. Southern California Occasional Papers in Linguistics 1, 71-95.

Matisoff, James A. 1978. Variational semantics in Tibeto-Burman: the 'organic' approach to linguistic comparison. Philadelphia: Institute for the Study of Human Issues.

Matisoff, James A. 1992. The mother of all morphemes: augmentatives and diminutives in areal and universal perspective. In: Martha Ratliff and Edith Schiller (eds.): Papers from the First Annual Meeting of the Southeast Asian Linguistics Society, 293-349. Tempe: Arizona State University, Program for Southeast Asian Studies.

Matisoff, James A. 2001. Genetic versus contact relationship: prosodic diffusibility in South-East Asian languages. In: Alexandra Y. Aikhenvald and R.M.W. Dixon (eds.): Areal diffusion and genetic inheritance: problems in comparative Linguistics, 291-327. Oxford: Oxford University Press.

Matisoff, James A. 2004. Areal semantics: is there such a thing? In: Anju Saxena (ed.): Himalayan languages, past and present, 347-393. Berlin: De Gruyter.

Matisoff, James A. 2008. The Tibeto-Burman reproductive system. Toward an etymological thesaurus. Berkeley/Los Angeles: University of California Press.

Mattei-Muller, Marie-Claude. 2007. Lengua y cultura Yanomamɨ. Diccionario ilustrado Yanomamɨ-Español / Español-Yanomamɨ. Caracas: Epsilon Libros.

Matteson, Esther. 1965. The Piro (Arawakan) language. Berkeley/Los Angeles: Berkeley University Press.

Matthews, P. H. 1974. Morphology. Cambridge: Cambridge University Press.

Mattissen, Johanna. 2003. Dependent-head synthesis in Nivkh. A contribution to a typology of polysynthesis. Amsterdam: Benjamins.

Mattissen, Johanna. 2004. Structural types of complex noun forms. In: Waldfried Premper (ed.): Dimensionen und Kontinua: Beiträge zu Hansjakob Seilers Universalienforschung, 35-56. Berlin: Akademieverlag.

Mazzocco, Michèle M. M. 1997. Children's interpretation of homonyms: a developmental study. Journal of Child language 24, 441-467.

McCarthy, Michael. 1999. What constitutes a basic vocabulary for spoken communication? Studies in English Language and Literature 1, 233-249.
McConvell, Patrick 2009a. Gurindji vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/31
McConvell, Patrick. 2009b. Loanwords in Gurindji, a Pama-Nyungan language of Australia. In: Haspelmath and Tadmor (eds.), 790-822.
McElhanon, K. A., and C. L. Voorhoeve. 1970. The Trans-New Guinea phylum. Explorations in deep-level genetic relationships. Canberra: Pacific Linguistics.
McRae, Ken, and Stephen Boisvert. 1998. Automatic semantic similarity priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 558-572.
McWhorter, John. 2001. The world's simplest grammars are Creole grammars. Linguistic Typology 5, 125-166.
Mead, Margaret. 1933. The Marsalai cult among the Arapesh, with special reference to the Rainbow Serpent beliefs of the Australian aboriginals. Oceania 4, 37-53.
Mel'čuk, Igor A. 1973. The structure of linguistic signs and possible formal-semantic relations between them. In Josette Rey-Debove (ed.): Recherches sur les systèmes signifiants, 103-135. The Hague: Mouton.
Mel'čuk, Igor A. 1976a. Das Wort zwischen Inhalt und Ausdruck, vol. 1. Ed. by Johann Biedermann. Munich: Wilhelm Fink.
Mel'čuk, Igor A. 1976b. On the notion of 'more/less compex semantically in derivation.' In: Mel'čuk 1976a, 115-137.
Mel'čuk, Igor A. 1976c. Zum Begriff der Wortbildung (Derivation). In: Mel'čuk 1976a, 63-88.
Mel'čuk, Igor A. 1981. Meaning-text models: a recent trend in Soviet linguistics. Annual Review of Anthropology 10, 27-62.
Mel'čuk, Igor. 2007. Lexical functions. In: Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn, and Neal R. Norrick (eds.): Phraseologie/Phraseology, 119-131. Berlin/New York: De Gruyter.
Mel'čuk, Igor A., and Alain Polguère. 1987. A formal lexicon in the meaning-text theory (or how to do lexica with words). Computational Linguistics 13, 261-275.
Menner, Robert J. 1936. The conflict of homonyms in English. Language 12, 229-244.
Messineo, Cristina. 2003. Lengua Toba (Guaycurú). Aspectos gramaticales y discursivos. Munich: Lincom Europa.
Mian Yan, Margaret. 2006. Introduction to Chinese dialectology. Munich: Lincom Europa.
Michelson, Karin. 1981. Three stories in Oneida. Ottawa: National Museums of Canada.
Michelson, Karin. 1990. The Oneida lexicon. Berkeley Linguistics Society 16, 73-85.
Michelson, Karin, and Mercy Doxtator. 2002. Oneida-English, English-Oneida Dictionary. Toronto: University of Toronto Press.
Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In: Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.): Language complexity: typology, contact, change, 23-41. Amsterdam: Benjamins.
Milewski, Tadeusz. 1950. La structure de la phrase dans les langues indigènes de l'Amérique du Nord. Lingua Posnaniensis 2, 162-207.

Miller, George A., Christiane Fellbaum, Judy Kegl, and Katherine J. Miller. 1988. Wordnet: an electronic lexical reference system based on theories of lexical memory. Princeton: Princeton University, Cognitive Science Laboratory.
Miller, Wick R. 1965. Acoma grammar and texts. Berkeley/Los Angeles: University of California Press.
Milner, George B. 1993. Samoan dictionary. Aotearoa: Polynesian Press.
Mithun, Marianne. 1984. How to avoid subordination. Berkeley Linguistics Society 10, 493-523.
Mithun, Marianne. 1999. The languages of Native North America. Cambridge: Cambridge University Press.
Mithun, Marianne. 2000. Noun and verb in Iroquoian languages: multicategorisation from multiple criteria. In: Petra M. Vogel and Bernard Comrie (eds.): Approaches to the typology of word classes, 397-420. Berlin: De Gruyter.
Mithun, Marianne. 2009. Polysynthesis in the Arctic. In: Marc-Antoine Mahieu and Nicole Tersis (eds.): Variations on polysynthesis: the Eskaleut Languages, 3-17. Amsterdam: Benjamins.
Mixco, Mauricio J. 1965. Kiliwa grammar. Dissertation, University of California at Berkeley.
Mixco, Mauricio J. 1977. The Kiliwa response to Hispanic culture. Berkeley Linguistics Society 3, 12-23.
Mixco, Mauricio J. 1985. Kiliwa dictionary. Salt Lake City: University of Utah.
Mixco, Mauricio J. 2000. Kiliwa. Munich: Lincom Europa.
Moerman, Daniel E. 1998. Native American ethnobotany. Portland: Timber Press.
Mondesir, Jones E. 1992. Dictionary of St. Lucian creole. Part 1: Kwéyòl-English, Part 2: English-Kwéyòl, ed. Lawrence D. Carrington. Berlin: De Gruyter.
Monod Bequelin, Aurore. 2006. Categories and compounding in Tzeltal. A preliminary approach. In: Ximena Lois and Valentina Vapnarsky (eds.): Lexical categories and root classes in Amerindian languages, 211-246. Bern: Peter Lang.
Montag, Susan. 2008. Diccionario Cashinahua. Tercera Edición.Yarinacocha, Peru: Instituto Lingüístico de Verano.
Moore, Bruce R. 1962. Correspondences in South Barbacoan Chibcha. In: Benjamin F. Elson (ed.): Studies in Ecuadorian Indian languages 1, 270-289. Norman: Summer Institute of Linguistics.
Moore, Bruce R. 1966. Diccionario Castellano-Colorado, Colorado-Castellano. Quito: Instituto lingüístico de Verano.
Moore, Bruce R. 1972. El sistema fonético del idioma Colorado. Boletín de Informaciones Científicas Nacionales 13, 72-78.
Morice, Adrien-Gabriel. 1932. The Carrier language. Mödling bei Wien/St. Gabriel: Verlag der Internationalen Zeitschrift 'Anthropos'.
Morris, Cliff. 1984. Tetun-English Dictionary. Canberra: Pacific Linguistics.
Morse, Nancy L., and Michael B. Maxwell. 1999. Cubeo grammar. Dallas: Summer Institute of Linguistics.
Morse, Nancy L., Jay K. Salser, and Neva Salser. 1999. Diccionario ilustrado bilingüe: Cubeo-Español, Español-Cubeo. Bogotá: Editorial Alberto Lleras Camargo.

Mosel, Ulrike, and Even Hovdhaugen. 1992. Samoan reference grammar. Oslo: Scandinavian University Press.
Moser, Rosmarie. 2004. Kabba. A Nilo-Saharan language of the Central African Republic. Munich: Lincom Europa.
Moser, Rosmarie, and Jean-Pierre Dingatoloum. 2001. Kabba-English-French dictionary. Munich: Lincom Europa.
Moss, Helen E., Ruth K. Ostrin, Lorraine K. Tyler, and William D. Marslen-Wilson. 1995. Accessing different types of lexical semantic information: evidence from priming. Journal of Experimental Psychology: Learning, Memory and Cognition 21, 863-883.
Mottin, Jean. 1978. Eléments de grammaire Hmong Blanc. Bangkok: Don Bosco Press.
Mous, Maarten. 2003. The linguistic properties of lexical manipulation and its relevance for Ma'á. In: Yaron Matras and Peter Bakker (eds.): The Mixed Language debate. Theoretical and empirical advance, 209-235. Berlin/New York: DeGruyter.
Mufwene, Salikoko S. 1980. Bantu class prefixes: inflectional or derivational? Chicago Linguistics Society 16, 246-258.
Mügge, Maximilian A. 1964. On truth and falsity in their ultramoral sense. In: Oscar Levy (ed.): The complete works of Friedrich Nietzsche. Vol. 2: early Greek philosophy, 171-192. New York: Russell & Russell. [Translation of Nietzsche 1873].
Munro, Pamela, and Catherine Willmond. 1994. Chickasaw: an analytical dictionary. Norman: University of Oklahoma Press.
Munro, Pamela, and Felipe H. Lopez. 1999. Di'csyonaary X:tèe'n Dìi'zh Sah Sann Lu'uc. Los Angeles: UCLA Chicano Studies Research Centre Publications.
Murdock, George Peter, and Douglas R. White. 1969. Standard cross-cultural sample. Ethnology 8, 329-369.
Murphy, Gregory L. 1988. Comprehending complex concepts. Cognitive Science 12, 529-562.
Murphy, Gregory L. 1996. On metaphoric representation. Cognition 60, 173-204.
Murphy, M. Lynne. 2003. Semantic relations and the lexicon. Cambridge: Cambridge University Press.
Muysken, Pieter. 2008. Introduction: conceptual and methodological issues in areal linguistics. In: Pieter Muysken (ed.): From linguistic areas to areal linguistics, 1-23. Amsterdam: Benjamins.
Najlis, Elena L. 1966. Lengua Abipona. Buenos Aires: Centro de Estudios Lingüistica, Universidad de Buenos Aires.
Nakayama, Toshihide. 2001. Nuuchahnulth (Nootka) Morphosyntax. Berkeley/Los Angeles: University of California Press.
Nakayama, Toshihide. 2003, ed. Caroline Little's Nuu-chah-nulth (Ahousaht) texts with grammatical analysis. Tokyo: ELPR.
Neely, James H. 1977. Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General 106, 226-254.

Neely, James H. 1991. Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: Derek Besner and Glyn W. Humphreys (eds.): Basic processes in reading, 264-336. Hillsdale: Lawrence Erlbaum.

Nettle, Daniel. 1995. Segmental inventory size, word length, and communicative efficiency. Linguistics 33, 359-367.

Nettle, Daniel. 1998. Coevolution of phonology and the lexicon in twelve languages of West Africa. Journal of Quantitative Linguistics 5, 240-245.

Nettle, Daniel. 1999. Linguistic diversity. Oxford: Oxford University Press.

Newman, Paul. 2000. The Hausa language: an encyclopedic reference grammar. New Haven: Yale University Press.

Newman, John. 2010. Review of Vanhove (2008). Language 86, 473-475.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. Language 62, 56-119.

Nichols, Johanna. 1992. Linguistic diversity in space and time. Chicago: Chicago University Press.

Nichols, Johanna. 1998. The origin and dispersal of languages: linguistic evidence. In: Nina Jablonski and Leslie C. Aiello (eds.): The origin and diversification of language, 127-70. San Francisco: California Academy of Sciences.

Nichols, Johanna. 2003. Diversity and stability in language. In: Joseph and Janda (eds.), 283-310.

Nichols, Johanna. 2009a. Linguistic complexity: a comprehensive definition and survey. In: Geoffrey Sampson, David Gil, and Peter Trudgill (eds.): Language complexity as an evolving variable, 110-125. Oxford: Oxford University Press.

Nichols, Johanna. 2009b. The great lexical shift: from verb-based to noun-based in late Proto-Slavic. Talk given at the Max Planck Institute for Evolutionary Anthropology, May 28, 2009.

Nichols, Johanna. 2010. Indeterminate motion verbs are denominal. In: Victoria Hasko and Renee Perlmutter (eds.): New approaches to Slavic verbs of motion, 47-66. Amsterdam: Benjamins.

Nichols, Johanna, and David A. Peterson. 1996. The Amerind personal pronouns. Langugage 72, 336-371.

Nichols, Johanna, and David A. Peterson. 2005. Personal pronouns: m-t and n-m patterns. In: Haspelmath et al. (eds.), 544-551.

Nichols, Johanna, and Balthasar Bickel. 2005a. Locus of marking in the clause. In: Haspelmath et al. (eds.), 98-101.

Nichols, Johanna, and Balthasar Bickel. 2005b. Locus of marking: whole-language typology. In: Haspelmath et al. (eds.), 106-109.

Nichols, Johanna, and Balthasar Bickel. 2005c. Possessive classification. In: Haspelmath et al. (eds.), 242-245.

Nichols, Lynn. 2008. Lexical semantic constraints on noun roots and noun borrowability. Studies in Language 32, 683-700.

Nichols, Lynn, and Johanna Nichols. 2007. Lexical derivational properties resist diffusion. Paper presented at the workshop on language contact and morphosyntactic variation and change, Seventh Biennial Conference of the Association of Linguistic Typology 2007, Paris.
Nies, Joyce. 1986. Diccionario Piro (Yine). Tokanchi gikshijikowaka-steno. Yarinacocha: Ministerio de Educación and Instituto Lingüístico de Verano.
Nietzsche, Friedrich. 2005/1873. Über Wahrheit und Lüge im außermoralischen Sinne. In: Giorgio Colli and Mazzino Montinari (eds.): Friedrich Nietzsche. Sämtliche Werke. Vol. 1, 873-890. Munich: dtv.
Nikolaeva, Irina. 2006. A historical dictionary of Yukaghir. Berlin: De Gruyter.
Nikolayev, Sergej L., and Sergej A. Starostin. 1994. A North Caucasian etymological dictionary. Moscow: Asterisk.
Nisbett, Richard E. 2003. The geography of thought: how Asians and Westerners think differently...and why. New York: Free Press.
Nisbett, Richard E., Kaiping Peng, Incheol Choi, and Ara Norenzayan. 2001. Culture and systems of thought: holistic versus analytic cognition. Psychological Review 108, 291-310.
Norenzayan, Ara, Edward E. Smith, Beom Jun Kim, and Richard E. Nisbett. 2002. Cultural preferenes for formal versus intuitive reasoning. Cognitive Science 26, 653-684.
Norenzayan, Ara, Incheol Choi, and Kaiping Peng. 2007. Perception and cognition. In: Shinobu Kitayama and Dov Cohen (eds.): Handbook of Cultural Psychology, 569-594. New York: Guilford Publications.
Nunberg, Geoffrey. 1979. The non-uniqueness of semantic solutions: polysemy. Linguistics and Philosophy 3, 143-184.
Ochoa Camargo, Gonçalo, 1997. Pequeno dicionário Bororo-Português. Campo Grande: Missão Salesiana de Mato Grosso, Universidade Católica Dom Bosco.
O'Grady, Geoffrey N. 1960. More on lexicostatistics. Current Anthropology 1, 338-339.
Öhmann, Emil. 1934. Über Homonymie und Homonyme im Deutschen. Helsinki: Annales Academiæ Scientarium Fennicæ, B XXXII.
O'Meara, Carolyn, and Jürgen Bohnemeyer. 2008. Complex landscape terms in Seri. Language Sciences 30, 316-339.
Orr, John. 1962. Three studies on homonymics. Edinburgh: Edinburgh University Press.
Osthoff, Hermann, and Karl Brugmann. 1878. Morphologische Untersuchungen auf dem Gebiete der Indogermansichen Sprachen. Vol. 1. Leipzig: Hirzel.
Oswalt, Robert L. 1960. A Kashaya grammar (Southwestern Pomo). Dissertation, University of California at Berkeley.
Oswalt, Robert L. 1975. K'ahshá:ya cahno Kalikakh. Working Paper No. 32, Kashaya Language in Culture Project, Department of Anthropology, California State College, Sonoma.
Oswalt, Robert L. n.d. Kashaya dictionary. Manuscript, Berkeley.
Overall, Simon. 2007. A grammar of Aguaruna. Dissertation, La Trobe University.
Packard, Jerome L. 1997a. Introduction. In: Packard (ed.), 1-34.

Packard, Jerome L. 1997b, ed. New approaches to Chinese word formation. Morphology, phonology and the lexicon in Modern and Ancient Chinese, 197-260. Berlin: De Gruyter.

Packard, Jerome L. 2000. The morphology of Chinese. A linguistic and cognitive approach. Cambridge: Cambridge University Press.

Panagl, Oswalt. 1976. Aspekte der kindersprachlichen Wortbildung. In: Gaberell Drachman (ed.): Akten der 3. Salzburger Jahrestagung für Linguistik, 79-100.

Paris, Catherine. 1989. West Circassian (Adyghe: Abzakh Dialect). In: George B. Hewitt (ed.): The indigenous languages of the Caucasus, vol. 2: The North West Caucasian languages, 155-260. Delmar/New York: Caravan Books.

Paris, Catherine, and Batouka, Níaz. 2005. Dictionnaire abzakh (tcherkesse occidental). Lovain/Paris/Dudley, MA: Peeters.

Parker, Gary J., and Amancio Chavez. 1976. Diccionario Quechua: Ancash-Huailas. Lima: Ministerio de Educación/Instituto de Estudios Peruanos.

Parks, Douglas R. 1976. A grammar of Pawnee. New York: Garland.

Parks, Douglas R., and Lula N. Pratt. 2008. A dictionary of Skiri Pawnee. Lincoln: University of Nebraska Press.

Parkvall, Mikael. 2000. Out of Africa: African Influences in Atlantic Creoles. London: Battlebridge.

Paul, Hermann. 1966/1880. Prinzipien der Sprachgeschichte. Tübingen: Niemeyer.

Pawley, Andrew. 1986. Lexicalization. In: Deborah Tannen and James E. Alatis (eds.): 1985 Georgetown University Roundtable on Languages and Linguistics, 98-120. Washington: Georgetown University Press.

Pawley, Andrew. 1993. A language which defies description by ordinary means. In: William A. Foley (ed.): The role of theory in language description, 87-129. Berlin: De Gruyter.

Pawley, Andrew. 1996a. Grammarian's lexicon, lexicographer's lexicon: worlds apart. In: Jan Svartvik (ed.): Words, 189-211. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.

Pawley, Andrew. 1996b. On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis. In: Janet Davidson, Geoffrey Irwin, Foss Leach, Andrew Pawley, and Dorothy Brown (eds.): Oceanic culture history: essays in honour of Roger Green, 387-410. Dunedin: New Zealand Journal of Archaeology.

Pawley, Andrew. 1999. Chasing rainbows. Implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In: Elizabeth Zeitoun and Paul Jen-kuei Li (eds.): Selected papers from the Eighth International Conference on Austronesian Linguistics, vol. 1, 95–138. Taipei: Academia Sinica.

Pawley, Andrew. 2006. On the size of the lexicon in preliterate languages: comparing dictionaries of Australian, Austronesian and Papuan languages. In: Josef Genzor (ed.): Favete linguis. Studies in honour of Viktor Krupa, 171-191. Bratislava: Institute of Oriental studies, Slovak Academy of Science.

Pawley, Andrew. 2009. The role of the Solomon Islands in the first settlement of Remote Oceania: bringing linguistic evidence to an archaeological debate. In: Karl Alexander Adelaar and Andrew Pawley (eds.): Austronesian historical linguistics and culture history. A festschrift for Robert Blust, 515-540. Canberra: Pacific Linguistics.

Pawley, Andrew, and Malcolm Ross. 1995. The prehistory of Oceanic languages: a current view. In: Peter Bellwood, James J. Fox and Darrell Tryon (eds.): The Austronesians: historical and comparative perspectives, 39-74. Canberra: Australian National University.

Payne, Doris L. 1987. Noun classification in the Western Amazon. Language Sciences 9, 21-44.

Payne, John R. 1990. Language universals and language types. In: Neville E. Collidge (ed.): An Encyclopedia of Language, 155-184. London: Longman.

Pederson, Eric. 1993. Geographic and manipulable space in two Tamil linguistic systems. In: Andrew U. Frank and Irene Campari (eds.): Spatial information theory, 294–311. Berlin: Springer.

Peirce, Charles Sanders. 1932. Speculative grammar. In: Charles Hartshorne and Paul Weiss (eds.): Collected papers of Charles Sanders Pierce, Vol. II: elements of Logic, 129-269. Cambridge, Mass.: Harvard University Press.

Perasso, José Antonio, and Diego Bracco. 1979. Vocabulario Lengua-Maskoy. Según un documento inedito de 1929, del General Jenaro Espinola. Asunción: Centro Paraguayo de Estudios Antropológicos.

Percival, W. Keith. 1988. A note on Thomas Hayne and his relation to Leibniz and Vico. New Vico Studies 6, 97-101.

Pericliev, Vladimir. 2004. There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. Linguistic Typology 8, 376-383.

Perkins, Revere D. 1989. Statistical techniques for determining language sample size. Studies in Language 13, 293-315.

Pfeifer, Wolfgang. 1993. Etymologisches Wörterbuch des Deutschen. Berlin: Akademieverlag.

Pillinger, Stephen Owen. 1989. Accent, tone and prosodic structure in Rendille. With particular reference to the nominal system. Dissertation, School of Oriental and Asian Studies, University of London.

Pillinger, Steve, and Letiwa Galboran. 1999. A Rendille dictionary, including a grammatical outline and an English-Rendille index. Cologne: Köppe.

Pitkin, Harvey. 1984. Wintu grammar. Berkeley/Los Angeles: University of California Press.

Pitkin, Harvey. 1985. Wintu dictionary. Berkeley/Los Angeles: University of California Press.

Plag, Ingo. 1999. Morphological productivity. Structural constraints in English derivation. Berlin: De Gruyter.

Plag, Ingo. 2003. Word-formation in English. Cambridge: Cambridge University Press.

Plank, Frans. 1981. Morphologische (Ir-)Regularitäten. Aspekte der Wortstrukturtheorie. Tübingen: Narr
Plank, Frans. 2003. There's more than one way to make sense of one-way implications – and sense they need to be made of. Linguistic Typology 7, 128-139.
Plomley, Norman James Brian. 1976. A word-list of the Tasmanian Aboriginal languages. Launceston: N.J.B. Plomley in association with the government of Tasmania.
Popovich, A. Harold, and Frances B. Popovich. 2005. Maxakalí-English dictionary; English-Maxakalí glossary. Cuiabá: Sociedade Internacional de Lingüística.
Poppe, Nicholas. 1954. Grammar of written Mongolian. Wiesbaden: Harrassowitz.
Pott, August Friedrich. 1860. Über Mannichfaltigkeit des sprachlichen Ausdrucks nach Laut und Begriff I, II, III. Zeitschrift für Völkerpsychologie und Sprachwissenschaft 1, 254-260, 345-348, 510-518.
Pott, August Friedrich. 1974/1884-1890. Einleitung in die Allgemeine Sprachwissenschaft. Amsterdam: Benjamins.
Powell, James V., and Fred Woodruff Sr. 1976. Quileute dictionary. Moscow, Idaho: Anthropological Research Notes Memoirs 3.
Pring, Julian T. 1982. The Oxford dictionary of Modern Greek. Oxford: Clarendon Press.
Proulx, Paul. 1988. Algic color terms. Anthropological Linguistics 30, 135-149.
Pukui, Mary Kawena, and Samuel H. Elbert. 1987. Hawaiian dictionary. Hawaiian-English, English-Hawaiian. Revised and Enlarged Edition. Honululu: University of Hawaii Press.
Pulleyblank, Edwin G. 1991. Lexicon of reconstructed pronunciation in early Middle Chinese, late Middle Chinese, and early Mandarin. Vancouver: University of British Columbia Press.
Pustejovsky, James. 1995. The generative lexicon. Cambrige, Ma./London: MIT Press.
Quine, W.V. 1973. The roots of reference. La Salle: Open Court.
R Development Core Team. 2009. R: a language and environment for statistical computing, version 2.9.2. Vienna: R Foundation for Statistical Computing.
Radden, Günter, and Zoltán Kövecses. 1999. Towards a theory of metonymy. In: Klaus-Uwe Panther and Günther Radden (eds.): Metonymy in language and thought, 17-59. Amsterdam: Benjamins.
Radden, Günter, and Klaus-Uwe Panther. 2004a. Introduction: reflections on motivation. In: Radden and Panther (eds.), 1-46.
Radden, Günter, and Klaus-Uwe Panther. 2004b, eds. Studies in linguistic motivation. Berlin: De Gruyter.
Ramamurti, G.V. 1931. A manual of the So:ra: (or Savara) language. Madras: Government Press.
Ramamurti, Rao Sahib G.V. 1938. Sora-English dictionary. Madras: Government Press.
Ramirez, Henri. 1994. Le parler Yanomami des Xamatauteri. Dissertation, Université de Aix-en-Provence.
Rankin, Robert L. 2003. The comparative method. In : Joseph and Janda (eds.), 183-212.
Räsänen, Martti. 1947. Regenbogen – Himmelsbrücke. Studia Orientalia 14, 3-11.

Ratliff, Martha. 1992. Form and function in tone languages. Berkeley Linguistics Society 18, 134-144.
Ratliff, Martha. 2009. White Hmong vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/25
Raun, Alo, and Andrus Saareste. 1965. Introduction to Estonian Linguistics. Wiesbaden: Harrassowitz.
Rayfield, Donald. 2002. Some distinctive characteristics of the vocabulary of Caucasian languages. In: Cruse et al. (eds.), 1039-1042.
Redéi, Károly. 1970. Die syrjänischen Lehnwörter im Wogulischen. The Hague: Mouton.
Rédei, Károly. 1988. Uralisches etymologisches Wörterbuch. 3 vols. Wiesbaden: Harrassowitz.
Reid, Aileen A., and Ruth G. Bishop. 1974. Diccionário Totonaco de Xicotepec de Juárez, Puebla. México, D.F.: Instituto Lingüístico de Verano.
Reid, Aileen A., and Sara Arjona de Watson. 1991. Gramática Totonaca de Xicotepec de Juárez, Puebla. Dallas: Summer Institute of Linguistics.
Renault-Lescure, Odile. 2009. Kali'na vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/38
Rensch, Karl H. 2002. Fiction, facts and insights. The perception of Polynesian languages in 18th and 19th century Europe. In: Kennosuke Ezawa (ed.): Linguistik jenseits des Strukturalismus: Akten des II. Ost-West-Kolloquiums Berlin 1998, 185-196. Tübingen: Narr.
Rettig, Wolfgang. 1981. Sprachliche Motivation. Zeichenrelationen von Lautform und Bedeutung am Beispiel französischer Lexikoneinheiten. Frankfurt: Peter Lang.
Rice, Sally. 2012. 'Our language is very literal.' Figurative expression in Dene Sųłiné [Athapaskan]. In: Anna Idström and Elisabeth Piirainen in cooperation with Tiber F.M. Falzett (eds.): Endangered metaphors, 21-76. Amsterdam: Benjamins.
Rich, Rolland G. 1999. Diccionario Arabela-Castellano. Lima: Instituto Lingüístico de Verano.
Riemer, Nick. 2002a. Interpreting semantic extension: metaphor and metonymy on different levels of lexical categorization. In: Cynthia Allen (ed.): Proceedings of the 2001 meeting of the Australian Linguistic Society. http://www.als.asn.au.
Riemer, Nick. 2002b. Remetonymizing metaphor: hypercategories in semantic extension. Cognitive Linguistics 12, 379-401.
Riemer, Nick. 2005. The semantics of polysemy: reading meaning in English and Warlpiri. Berlin: De Gruyter.
Riemer, Nick. 2010. Introducing semantics. Cambridge: Cambridge University Press.
Rießler, Michael. 2009. Kildin Saami vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/14.
Rigby, Nora, and Robin Schneider. 1989. Dictionary of the Rama language: Rama-English-Rama-Creole-Spanish/English-Rama. Berlin: Reimer.
Rijkhoff, Jan. 2009. On the (un)suitability of semantic categories. Linguistic Typology 13, 95-104.
Rijkhoff, Jan, and Dik Bakker. 1998. Language sampling. Linguistic Typology 2,263-314.

Rijkhoff, Jan, Dik Bakker, Kees Hengeveld, and Peter Kahrel. 1993. A method of language sampling. Studies in Language 17, 169-203.
Rix, Helmut. 2002. Wurzeletymologie. In: Cruse et al. (eds.), 1333-1339.
Rix, Helmut, and Martin Kümmel. 2001. Lexikon der Indogermanischen Verben. Second edition. Wiesbaden: Harrassowitz.
Robert, Stéphane. 1997. Variation des représentations linguistiques: des unités à l'énoncé. In: Catherine Fuchs and Stéphane Robert (eds.): Diversité des langues et représentations cognitives. Paris: Ophrys, 25-39.
Robinson, Lila Wistrand, and James Armagost. 1994. Comanche dictionary and grammar. Dallas: Summer Institute of Linguistics.
Rodrigues, Aryon Dall'Igna. 1964. A classificação do tronco lingüístico Tupí. Revista de Antropologia 12, 99-104.
Romaine, Suzanne. 1988. Pidgin and Creole languages. London: Longman.
Romaine, Suzanne. 2002. Lexical structure in Pidgins and Creoles. In: Cruse et al. (eds.), 1092-1095.
Rood, David S. 1976. Elementary bilingual dictionary English-Lakhóta, Lakhóta-English. Boulder: Department of Linguistics, University of Colorado.
Rosch, Eleanor, and Carolyn B. Mervis. 1975. Family resemblances: studies in the internal structure of categories. Cognitive Psychology 7, 573-605.
Rosch, Eleanor. 1978. Principles of categorization. In: Eleanor Rosch and Barbara B. Lloyd (eds.): Cognition and categorization, 27-48. Hillsdale: Lawrence Erlbaum.
Ross, Malcolm D. 1980. Some elements of Vanimo, a New Guinea tone language. Papers in New Guinea Linguistics 20, 77-109.
Ross, M.D. 1988. Proto Oceanic and the Austronesian languages of Western Melanesia. Canberra: Pacific Linguistics.
Ross, Malcolm D. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In: Durie and Ross (eds.), 180-217.
Ross, Malcolm D. 2002. Takia. In: Lynch et al. (eds.), 216-248.
Ross, Malcolm. 2009. Takia vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/29
Rude, Noel Emerson. 1985. Studies in Nez Perce grammar and discourse. Dissertation, University of Oregon.
Rudes, Blair A. 1999. Tuscarora-English, English-Tuscarora dictionary. Toronto: University of Toronto Press.
Ruhlen, Merrit. 1987. A guide to the world's languages. Volume 1: classification. Stanford: Stanford University Press.
Sadovszky, Otto J. 1973. The reconstruction of IE **Pisko* and the extension of its semantic sphere. Journal of Indo-European Studies 1, 81-100.
Sakhno, Serguei. 1999. Typologie des parallèles lexicaux russes-français dans une perspective sémantico-historique. Slovo 22/23, 287-313.
Salamanca, Daniel. 1988. Elementos del gramatica de Miskito. Dissertation, Massachusetts Institute of Technology.

Sánchez, Francisco. 1995/1587. Minerva o de causis linguae Latinae. Ed. by Eustaquio Sanchez Salor. Cáceres: Instituto Cultural 'El Brocense', Universidad de Extremadura.
Sandra Dominiek. 1998. What linguists can and can't tell you about the human mind: a reply to Croft. Cognitive Linguistics 9, 361-378.
Santa Ynez Band of Chumash Indians in collaboration with Richard Applegate. 2008. Samala-English Dictionary.
Sapir, Edward. 1911. Some aspects of Nootka language and culture. American Anthropologist 13, 15-28.
Sapir, Edward. 1912. Language and environment. American Anthropologist 14, 226-242.
Sapir, Edward. 1916/1949. Time perspective in Aboriginal American culture. A study in method. In: David G. Mandelbaum (ed.): Selected writings of Edward Sapir in language, culture and personality, 389-462. Berkely/Los Angeles: University of California Press.
Sapir, Edward. 1921/1970. Language. An introduction to the study of speech. London: Hart-Davis.
Sapir, Edward. 2002. The psychology of culture. A course of lectures. Reconstructed and ed. by Judith T. Irvine. Berlin: De Gruyter.
Sapir, Edward, and Morris Swadesh. 1960. Yana dictionary. Berkeley/Los Angeles: University of California Press.
Sara, Salomon. 2001. A tri-lingual dictionary of Emberá-Spanish-English. Munich: Lincom Europa.
Sarton, George. 1959. A history of science: hellenistic science and culture in the last three centuries B.C. Cambridge, MA: Harvard University Press.
Sasse, Hans-Jürgen. 1985. Sprachkontakt und Sprachwandel. Die Gräzisierung der albanischen Mundarten Griechenlands. Papiere zur Linguistik 32, 37-95.
Sasse, Hans-Jürgen. 1991. Predication and sentence constitution in universal perspective. In: Zaefferer (ed.), 75-95.
Sasse, Hans-Jürgen. 1993. Das Nomen - eine universelle Kategorie? Sprachtypologie und Universalienforschung 46, 187-221.
Sasse, Hans-Jürgen. 2001. Scales between nouniness and verbiness. In: Haspelmath et al. (eds.), 495-509.
Sasse, Hans-Jürgen. 2002. Some distinctive characteristics of the vocabulary of Cushitic languages. In: Cruse et al. (eds.), 1053-1060.
Saussure, Ferdinand de. 1916/1967. Cours de linguistique générale. Paris: Payot.
Saveleva, Valentina N., and Cuner M. Taksami. 1965. Russko-Nivchskij slovar'. 17300 slov. Moskva: Izdat. Sovetskaja Enciklopedija.
Saveleva, Valentina N., and Cuner M.Taksami. 1970. Nivchsko-Russkij slovar'. svyse 13000 slov. Moskva: Izdat. Sovetskaja Enciklopedija.
Sawyer, Jesse O. 1965. English-Wappo vocabulary. Berkeley/Los Angeles: University of California Press.
Sawyer, Jesse O., and Alice Schlichter. 1984. Yuki vocabulary. Berkeley/Los Angeles: University of California Press.

Schachter, Paul. 1985. Parts-of-speech Systems. In: Shopen (ed.), 3-61.
Schadeberg, Thilo. 2009. Swahili vocabulary. In: Haspelmath and Tadmor (eds.). http://wold.livingsources.org/vocabulary/1.
Schebeck, Bernard. 1978. Names of body-parts in North-East Arnhem Land. In: Lester Richard Hiatt (ed.): Australian aboriginal concepts, 168-177. Canberra: Australian Institute of Aboriginal Studies.
Scheidegger, Jean. 1981. Arbitraire et motivation en francais et en allemand. Examen critique des thèses de Charles Bally. Bern: Francke.
Schieffelin, Bambi B., and Steven Feld. 1998. Bosavi-English-Tok Pisin dictionary (Papua New Guinea). Canberra: Australian National University.
Schindler, Jochem. 1969. Die idg. Wörter für 'Vogel' und 'Ei'. Die Sprache 15, 144-167.
Schmidt, Christopher K. 2009. Japanese vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/21.
Schmidt, Günter Dietrich. 1982. Kann ein 'elementares Sprachzeichen' polysem sein? Linguistische Berichte 79, 1-11.
Schmidt, Hans. 2003. Loanword strata in Rotuman. In: Henning Andersen (ed.): Language contacts in prehistory: studies in stratigraphy, 201-240. Amsterdam: Benjamins.
Schmidt, Wilhelm. 1952. Die Tasmanischen Sprachen. Utrecht: Spectrum.
Schröpfer, Johannes. 1974a. The purpose of a comparative dictionary of semantic change (a dictionary of semantic parallels). Semantische Hefte 1, 45-49.
Schröpfer, Johannes. 1974b. Vorwort. Semantische Hefte 1, 4-12.
Schröpfer, Johannes. 1979-1994. Wörterbuch der vergleichenden Bezeichungslehre. Heidelberg: Winter.
Schröpfer, Johannes. 1982. Das Wörterbuch der vergleichenden Bezeichnungslehre als Instrument der Slavistik und der allgemeinen Sprachwissenschaft. In: Eberhard Reißner (ed.): Literatur- und Sprachentwicklung in Osteuropa im 20. Jahrhundert, 157-165. Berlin: Berlin-Verlag.
Schultze-Berndt, Eva. 2000. Simple and complex verbs in Jaminjung. A study of event categorisation in an Australian language. Nijmegen: MPI Series in Psycholinguistics.
Schütz, Albert J. 1985. The Fijian language. Honolulu: University of Hawaii Press.
Searle, John R. 1969/1980. Speech Acts: an essay in the philosophy of language. Cambridge: Cambridge University Press.
Segert, Stanislav. 1991. The use of comparative Semitic material in Hebrew lexicography. In: Alan S. Kaye (ed.): Semitic studies. In honor of Wolf Leslau on the occasion of his 85th birthday, 1426-1434. Wiesbaden: Harrassowitz.
Seifart, Frank. 2005. The structure and use of shape-based noun classes in Miraña (North West Amazon). Nijmegen: Max Planck Institute for Psycholinguistics.
Seifart, Frank, and Doris L. Payne. 2007. Nominal classification in the North West Amazon: issues in areal diffusion and typological characterization. International Journal of American Linguistics 73, 381-387.

Seifart, Frank, and Julien Meyer. 2010. Bora drum communication, the typology of emulated speech, and prosodic typology. Paper given at the Arbeitsgruppe on prosodic typology: state of the art and future prospects at the annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Berlin, February 24-26 2010.

Seiler, Hansjakob. 1975. Die Prinzipien der deskriptiven und der etikettierenden Benennung. In: Hansjakob Seiler (ed.): Linguistic workshop III. Arbeiten des Kölner Universalienprojekts 1974, 2-57. München: Fink.

Seiler, Hansjakob. 1976. Introductory notes to a grammar of Cahuilla. Cologne: Arbeiten des Kölner Universalien-Projekts (akup) 20.

Seiler, Hansjakob. 1977. Cahuilla grammar. Banning: Malki Museum Press.

Seiler, Hansjakob, and Kojiro Hioki. 1979. Cahuilla dictionary. Banning: Malki Museum Press.

Semenza, Carlo, and Sara Mondini. 2006. The neuropsychology of compound words. In: Libben and Jarema (eds.), 71-95.

Seto, Ken-ichi. 2003. Metonymic polysemy and its place in meaning extension. In: Brigitte Nerlich, Zazie Todd, Vimala Herman, and David D. Clarke (eds.): Polysemy: flexible patterns of meaning in mind and language, 195-214. Berlin: De Gruyter.

Shepherd, John W., Graham M. Davies, and Hadyn D. Ellis. 1981. Studies of cue saliency. In: Graham M. Davies, Hadyn D. Ellis, and John W. Shepherd (eds.): Perceiving and remembering faces, 105-132. London: Academic Press.

Sherzer, Joel. 1973. Areal linguistics in North America. In: Thomas A. Sebeok (ed.): Current trends in linguistics. Vol. 10.2: Linguistics in North America, 749-795. The Hague: Mouton.

Shi, Yuzhi. 2002. The establishment of Modern Chinese grammar. The formation of the resultative construction and its effects. Amsterdam: Benjamins.

Shields, Lynne W., and David A. Balota. 1991. Repetition and associative context effects in speech production. Language and Speech 34, 47-55.

Shopen, Timothy. 1985, ed. Language typology and syntactic description. Vol. I: Clause structure. Cambridge: Cambridge University Press.

Shopen, Timothy. 2007, ed. Language typology and syntactic description. Vol. III: Grammatical categories and the lexicon. Second Edition. Cambridge: Cambridge University Press.

Siewierska, Anna. 2005. Verbal person marking. In: Haspelmath et al. (eds.), 414-417.

Silva Noelli, Francisco. 2008. The Tupi expansion. In: Helaine Silverman and William Harris Isbell (eds.): Handbook of South American archaeology, 659-670. New York: Springer.

Silverman, Daniel. 1997. Tone sandhi in Comaltepec Chinantec. Language 73, 473-492.

Silverman, Daniel. 2010. Neutralization and anti-homophony in Korean. Journal of Lingusitics 46, 453-482.

Simons, Gary F. 1982. Word taboo and comparative Austronesian linguistics. In: Amran Halim, Lois Carrington, and S.A. Wurm (eds.): Papers from the Third International Conference on Austronesian Linguistics, Vol. 3: Accent on variety, 157-226. Canberra: Australian National University.

Sloman, Steven A. 1996. The empirical case for two systems of reasoning. Psychological Bulletin 119, 3-22

Smith, Barry, and David M. Mark. 2001. Geographical categories: an ontological investigation. International Journal of Geographical Information Science 15, 591-612.

Smith, Barry, and David M. Mark. 2003. Do mountains exist? Towards an ontology of landforms. Environment and Planning B: Planning and Design 30, 411-427.

Smith, Edward E., Daniel N. Osherson, Lance J. Rips, and Margaret Keane. 1988. Combining prototypes: a selective modification model. Cognitive Science 12, 485-527.

Smith, Kenneth D. 1979. Sedang grammar: phonological and syntactic structure. Canberra: Australian National University.

Smith, Kenneth D. 2000. Sedang dictionary. With English, Vietnamese, and French glossaries. Salaya: Mahidol University & Dallas: Summer Institute of Linguistics.

Smith-Stark, Thomas C. 1994. Mesoamerican calques. In: Carolyn J. MacKay and Verónica Vázquez (eds.): Investigaciones lingüisticas en Mesoamérica, 15-50. Mexico: Universidad Nacional Autónoma de México.

Song, Jae Jung. 2005. Nonperiphrastic causative constructions. In: Haspelmath et al. (eds.), 450-453.

Song, Jae Jung. 2010, ed. The Oxford handbook of linguistic typology. Oxford: Oxford University Press.

Spencer, Andrew. 1995. Incorporation in Chukchi. Language 71, 439-489.

Stanhope, John M. 1980. The language of the Rao People, Grengabu, Madang province, Papua New Guinea. Canberra: Australian National University.

Starosta, Stanley. 1992. Sora combining forms and pseudo-compounding. Mon Khmer Studies 18/19, 77-105.

Stassen, Leon. 2005a. Predicative adjectives. In: Haspelmath et al. (eds.), 478-481.

Stassen, Leon. 2005b. Zero copula for predicate nominals. In: Haspelmath et al. (eds.), 486-489.

Stebbins, Tonya. 2005. Nominal classification in Mali. Anthropological Linguistics 47, 77-131.

Stebbins, Tonya. n.d. Formulaicity in language description: formulaic sequences in Mali. Manuscript.

Stebbins, Tonya N., with the assistance of J. Tayul. In press. Mali (Baining) dictionary. Canberra: Australian National University.

Stevenson, Roland. n.d. Katcha dictionary. Manuscript.

Stewart, Pamela, and Andrew J. Strathern. 1997. Netbags revisited: cultural narratives from Papua New Guinea. Pacific Studies 20, 1-30.

Stonham, John T. 2005. A concise dictionary of the Nuuchahnulth language of Vancouver Island. Lewiston: Edwin Mellen Press.

Story, Gillian L. 1984. Babine & Carrier phonology. A historically oriented study. Dallas: Summer Institute of Linguistics.

Street, John C. 1963. Khalkha structure. Bloomington: Indiana University & The Hague: Mouton.

Suárez, Maria Matilde. 1977. La lengua Sáliva. Caracas: Centro de Lenguas Indígenas, Instituto de Investigaciones Históricas, Univ. Católica Andrés Bello.
Sullivan, Karen, and Eve Sweetser. 2010. Is 'generic is specific' a metaphor? In: Faye Perrill, Vera Tobin, and Mark Turner (eds.): Meaning, form, and body, 309-328. Stanford/Chicago: CSLI/University of Chicago Press.
Svanlund, Jan. 2007. Metaphor and convention. Cognitive Linguistics 18, 47-89.
Swadesh, Morris. 1971. The origin and diversification of language. Ed. by Joel Sherzer. Chicago: Aldine.
Sweetser, Eve. 1990. From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure. Cambridge: Cambridge University Press.
Szemerényi, Oswald J. L. 1990. Introduction to Indo-European linguistics. Fourth edition. Oxford: Oxford University Press.
Tadmor, Uri. 2009. Loanwords in the world's languages: findings and results. In: Haspelmath and Tadmor (eds.), 55-75.
Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. Diachronica 27, 226-246.
Tagliavini, Carlo. 1949/1982. Di alcuni denominazioni della 'pupilla' (studio di onomasiologia, con speciale riguardo alle lingue camito-semitiche e negro-africane) [On some denominations of the 'pupil' (an onomasiological study, with special reference to the Hamito-Semitic and Black African languages]. In: Carlo Tagliavini: Scritti Minori, 529-568. Bologna: Patron.
Talmy, Leonard. 1988/2006. The relation of grammar to cognition. In: Dirk Geeraerts (ed.): Cognitive linguistics: basic readings, 69-108. Berlin: De Gruyter. Originally published in Brygida Rudzka-Ostyn (ed.): Topics in Cognitive Linguistics, 165-205. Amsterdam: Benjamins.
Talmy, Leonard. 2000. Toward a cognitive semantics. Vol. 1. Cambridge, Ma: MIT Press.
Talmy, Leonard. 2007. Lexical typologies. In: Shopen (ed.), 66-168.
Tappolet, Ernst. 1895. Die romanischen Verwandtschaftsnamen: mit besonderer Berücksichtigung der französischen und italienischen Mundarten; ein Beitrag zur vergleichenden Lexikologie. Strassburg: Trübner.
Taylor, John R. 2002. Cognitive grammar. Oxford: Oxford University Press.
Taylor, John R. 2003. Linguistic categorization. Oxford: Oxford University Press.
Terrill, Angela. 1996-2003. Unpublished Lavukaleve field notes.
Terrill, Angela. 1999. A grammar of Lavukaleve: a Papuan language of the Solomon islands. Canberra: Australian National University.
Terrill, Angela. 2002. Systems of nominal classification in East Papuan languages. Oceanic Linguistics 41, 63-88.
Terrill, Angela 2006. Body part terms in Lavukaleve, a Papuan language of the Solomon Islands. Language Sciences 28, 304-322.
Thayer, James Edward. 1978. The deep structure of the sentence in Sara-Ngambay dialogues. Dallas: Summer Institute of Linguistics.
The Oxford English Dictionary. Second Edition. 1989. Oxford: Oxford University Press.

Thiesen, Wesley. 1996. Gramática del idioma Bora. Pucallpa: Instituto Lingüístico de Verano.
Thiesen, Wesley, and Eva Thiesen. 2008. Diccionario Bora-Castellano, Castellano-Bora. Segunda Edición. Lima: Instituto Lingüístico de Verano.
Thomason, Sarah Grey, and Terrence Kaufman. 1988. Language contact, creolization and genetic linguistics. Berkeley/Los Angeles: University of California Press.
Thompson, Laurence C. 1991. A Vietnamese reference grammar. Honolulu: University of Hawai'i Press.
Thompson, Sandra A., Joseph Sung-Yul Park, and Charles N. Li. 2006. A reference grammar of Wappo. Berkeley/Los Angeles: University of California Press.
Thorne, David A. 1993. A comprehensive Welsh grammar. Gramadeg Cymraeg cynhwysfawr. Oxford: Blackwell.
Thurston, William R. 1989. How exoteric languages build a lexicon: esoterogeny in West New Britain. In: Ray Harlow and Robin Hooper (eds.): VICAL 1: Oceanic Languages. Papers from the Fifth International Conference on Austronesian Linguistics, 555-579. Auckland: Linguistic Society of New Zealand.
Trager, George L. 1939. 'Cottonwood' = 'tree'. A south-western linguistic trait. International Journal of American Linguistics 9, 117-118.
Trask, Robert L. 1997. The history of Basque. London: Routledge.
Traugott, Elizabeth Closs. 2010. Review of Vanhove (2008). Linguistic Typology 14, 292-299.
Traugott, Elizabeth Closs, and Richard B. Dasher. 2002. Regularity in semantic change. Cambridge: Cambridge University Press.
Trudgill, Peter. 1996. Dialect typology: isolation, social network and phonological structure. In: Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh (eds.): Towards a social science of language, Vol. 1: Variation and change in language and society, 3-22. Amsterdam: Benjamins.
Trudgill, Peter. 2002. Linguistic and social typology. In: J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.): The handbook of language variation and change, 707-728. Malden: Blackwell.
Trudgill, Peter. 2004. Linguistic and social typology. The Austronesian migrations and phoneme inventories. Linguistic Typology 8, 305-320.
Tucker, Archibald N., Margaret A. Bryan, and Wolf Leslau. 1966. Linguistic analyses: the Non-Bantu languages of North-Eastern Africa, with a supplement on the Ethiopic languages by Wolf Leslau. Oxford: Oxford University Press.
Tuggy, David. 1987. Scarecrow nouns, generalizations, and Cognitive Grammar. In: Scott de Lancey and Russell S. Tomlin (eds.): Proceedings of the Third Annual Meeting of the Pacific Linguistics Conference, 307-320. Eugene: Department of Linguistics, University of Oregon.
Tuggy, David. 1993. Ambiguity, polysemy, and vagueness. Cognitive Linguistics 4, 273-290.
Tuggy, David. 2005. Cognitive approach to word-formation. In: Pavol Štekauer and Rochelle Lieber (eds.): Handbook of word-formation, 233-265. Dordrecht: Springer.

Tuite, Kevin. The myth of the Caucasian Sprachbund: The case of ergativity. Lingua 108, 1-29.
Turner, Nancy J., Marianne Boelscher Ignace, and Brian D. Compton. 1998. Secwepemc (Shuswap) tree names: key to the past? In: Ewa Czaykowski-Higgins and Marvin Dale Kinkade (eds.): Salish languages and linguistics: theoretical and descriptive perspectives, 387-417. Berlin: De Gruyter.
Turner, Nancy J., and Cecil H. Brown. 2004. Grass, hay, low growth and weeds. Utility and semantics of Interior Salish botanical terms. In: Donna B. Gerdts and Lisa Matthewson (eds.): Studies in Salish linguistics in honor of M. Dale Kinkade, 410-428. Missoula: University of Montana Occasional Papers in Linguistics 17.
Turner, Paul. 1966. Highland Chontal grammar. Dissertation, University of Chicago.
Turner, Paul, and Shirley Turner. 1971. Chontal to Spanish-English, Spanish to Chontal dictionary. Tucson: The University of Arizona Press.
Turner, R. L. 1966. A comparative dictionary of the Indo-Aryan languages. New Delhi: Motilal Banarsidass Publishers.
Tylor, Edward Burnett. 1871. Primitive culture. Researches into the development of mythology, philosophy, religion, language, art, and custom. London: John Murray.
Ullmann, Stephen. 1962. Semantics. An introduction to the science of meaning. Oxford: Blackwell.
Ullmann, Stephen. 1966. Semantic universals. In: Joseph H. Greenberg (ed.): Universals of language. Report of a conference held at Dobbs Ferry, New York, April 13-15, 1961, 217-262. Cambridge, Mass./London: MIT Press.
Ultan, Russell. 1975. Descriptivity grading of Finnish body-part terms. Cologne: Arbeiten des Kölner Universalien-Projekts 16.
Ultan, Russell. 1976. Descriptivity in the domain of body-part terms. Cologne: Arbeiten des Kölner Universalienprojekts Nr. 21.
Umbreit, Birgit. 2010. Does *love* come from *to love* or *to love* from *love*? Why lexical motivation has to be regarded as bidirectional. In: Alexander Onysko and Sascha Michel (eds.): Cognitive perspectives on word formation, 301-333. Berlin: De Gruyter.
Ungerer, Friedrich. 2002. Derivational word-formation in English. Anglia 120, 534-567.
Ungerer, Friedrich. 2007. Derivational morphology and word-formation. In: Dirk Geeraerts and Hubert Cuyckens (eds.): The Oxford handbook of Cognitive Linguistics, 991-1025. Oxford: Oxford University Press.
University Press PLC. n.d. A dictionary of the Yoruba language. Ibadan: University Press Nigeria.
Urban, Matthias. 2008. Deskriptivität als Subtyp lexikalischer Motivation – Eine quantitative Studie in Synchronie und Diachronie. Arbeitspapier Nr. 54 (Neue Folge). Cologne: Institute for Linguistics, University of Cologne.
Urban, Matthias. 2009. 'Sun' and 'Moon' in the Circum-Pacific language area. Anthropological Linguistics 51, 328-346.
Urban, Matthias. 2010. 'Sun' = 'eye of the day': a Linguistic pattern of Southeast Asia and Oceania. Oceanic Linguistics 49, 568-579.

Urban, Matthias. 2011. Asymmetries in overt marking and directionality in semanic change. Journal of Historical Linguistics 1, 3-47.
Urban, Matthias. Forthcoming. Cultural explanations for recurrent lexico-semantic associations: the pupil of the eye revisited. Proceedings of the Young Linguists' Meeting in Poznań 2010. Poznán: Adam Mickiewicz University.
Van den Berg, René, and La Ode Sidu. 1996. Muna - English dictionary. Leiden: KITLV Press.
Van Engelenhoven, Aone, and Catharina Williams-van Klinken. 2005. Tetun and Leti. In: K. Alexander Adelaar and Nikolaus Himmelmann (eds.): The Austronesian languages of Asia and Madagascar, 735-768. London: Routledge.
Van Klinken, Catharina L. 1999. A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor. Canberra: Pacific Lingustics.
Vajda, Edward J. 2004a. Distinguishing between referential and grammatical function in morphological typology. In: Zygmunt Frajzyngier, Adam Hodges, and David S. Rood (eds.): Linguistic diversity and language theories, 399-422. Amsterdam: Benjamins.
Vajda, Edward J. 2004b. Ket. Munich: Lincom Europa.
Vajda, Edward, and Andrey Nefedov. 2009. Ket vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/18.
Vamarasi, Marit. 2002. Rotuman. Munich: Lincom Europa.
Van der Auwera, Johan, Nina Dobrushina, and Valentin Goussev. 2005. Imperative/hortative Systems. In: Haspelmath et al. (eds.), 294-297.
Vanhove, Martine. 2008, ed. From polysemy to semantic change. Amsterdam: Benjamins.
Vico, Giambattista. 1744/1976. La scienza nouva e altri scritti. Ed. by Nicola Abbagnano. Torino: UTET.
Vico, Giambattista. 1735/1843. Principj di una scienza nouva d'intorno alla comune natura della nazioni. Milan: Societa Tipografica de Classici Italiani.
Vidal, Alejandra, and Verónica Nercesian. 2009. Wichí vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/40.
Viñas Urquiza, María Teresa. 1974. Lengua Mataca. Buenos Aires: Centro de Estudios Lingüísticos, Facultad de Filosofía y Letras, Universidad de Buenos Aires.
Visser, L.E., and C.L. Voorhoeve. 1987. Sahu-Indonesian-English dictionary and grammar sketch. Dordrecht: Foris Publications.
Vittadello, Alberto. 1988. Cha'palaachi. El idioma Cayapa. Esmeraldas: Pontificia Universidad Católica del Ecuador, Sede de Esmeraldas.
Voegelin, C. F., and Dell H. Hymes. 1953. A sample of North American Indian dictionaries with reference to acculturation. Proceedings of the American Philosophical Society 97, 634-644.
Voegelin, Charles F., and Florence M. Voegelin. 1977. Classification and index of the world's languages. New York: Elsevier.
Vogel, Alan. 2006. Jarawara-English dictionary. Cuiabá: Sociedade Internacional de Lingüística.
Wälchli, Bernhard. 2005. Co-compounds and natural coordination. Oxford: Oxford University Press.

Walter, Heribert. 1976. Zum Problem der Deskriptivität am Beispiel deutscher Verbalderivation. Cologne: Arbeiten des Kölner Universalien-Projekts 26.
Waltereit, Richard. 1999. Grammatical constraints of metonymy. On the role of the direct object. In: Klaus-Uwe Panther and Günther Radden (eds.): Metonymy in language and thought, 233-254. Amsterdam: Benjamins.
Warman, Adolfo I. 1959. Diccionario trilingue: Miskito-Español-Inglés; Español-Miskito, Inglés-Miskito. Managua: Talleres Nacionales.
Warren, Beatrice. 1978. Semantic patterns of noun-noun compounds. Göteborg: Acta Universitatis Gothoburgensis.
Warren, Howard Crosby. 1921. A history of the association psychology. London: Constable.
Watkins, Laurel J. 1984. A grammar of Kiowa. Lincoln: University of Nebraska Press.
Wehr, Hans. 1976. A dictionary of Modern Written Arabic. Ed. by J. Milton Cowan. Third edition. Ithaca: Spoken Language Services.
Weinreich, Uriel. 1953. Languages in contact. Problems and findings. The Hague: Mouton.
Weisberger, Leo. 1971. Die geistige Seite der Sprache und ihre Erforschung. Düsseldorf: Schwann.
Weldon, Mary, S. 1991. Mechanisms underlying priming on perceptual tests. Journal of Experimental Psychology: Learning, Memory, and Cognition 17, 526-541.
Welmers, William E. 1968. Efik. Ibadan: University of Ibadan, Institute of African Studies.
Werner, Heinrich. 1997. Die Ketische Sprache. Wiesbaden: Harrassowitz.
Werner, Heinrich. 1998. Probleme der Wortbildung in den Jenissej-Sprachen. Munich: Lincom Europa.
Westrum, Peter, and Susan Westrum. 1986. Ol unggwanfer Berik olem. Perbendaharaan kata bahasa Berik. Berik vocabulary. Jayapura: Universitas Cenderawasih and Summer Institute of Linguistics.
Widmann, Thomas, and Peter Bakker. 2006. Does sampling matter? A test in replicability, concerning numerals. Linguistic Typology 10, 83-96.
Wiebusch, Thekla. 2009. Mandarin Chinese vocabulary. In: Haspelmath and Tadmor (eds.), http://wold.livingsources.org/vocabulary/22
Wiesemann, Ursula. 1972. Die phonologische und grammatische Struktur der Kaingáng-Sprache. The Hague/Paris: Mouton.
Wiesemann, Ursula Gojtéj. 2011. Dicionário Kaingang-Português, Português-Kaingang. Curitiba: Editora Evangélica Esperança.
Wilbur, Ronnie B., and Lise Menn. 1975. Towards a redefinition of psychological reality: on the internal structure of the lexicon. San Jose Occasional Papers in Linguistics 1, 212-221.
Wilkins, David P. 1996. Natural tendencies of semantic change and the search for cognates. In: Durie and Ross (eds.), 264-304.
Willems, Dominique. 1983. Syntax and semantics: on the search of constants in verbal polysemy. In: Shiro Hattori and Kazuko Inoue (eds.): Proceedings of the XIIIth International Congress of Linguists, August 29 - September 4 1982, 425-429. Tokyo: CIPL.

Williams, Edna Rees. 1944. The conflict of homonyms in English. New Haven: Yale University Press.
Williams, Stephen J. 1980. A Welsh grammar. Cardiff: University of Wales Press.
Williams-Van Klinken, Catharina L., John Hayek, and Rachel Nordlinger. 2002. Tetun Dili: a grammar of an East Timorese language. Canberra: Australian National University.
Winter, Bodo. 2011. A very basic tutorial for performing linear mixed effects analyses. With special reference to psycholinguistic experiments. http://www.bodowinter.com/tutorial/bw_LME_tutorial.pdf
Winter, Werner. 1992. Borrowing and non-Borrowing in Walapai. In: Ernst Håkon Jahr (ed.): Language Contact. Theoretical and empirical studies, 213- 228. Berlin: De Gruyter.
Wipio Deicat, Gerardo. 1996. Diccionario Aguaruna-Castellano, Castellano-Aguaruna. Lima: Instituto Lingüístico de Verano.
Wise, Mary Ruth. 1999. Small language families and isolates in Peru. In: Dixon and Aikhenvald (eds.), 307-340.
Wisniewski, Edward J. 1996. Construal and similarity in conceptual combination. Journal of Memory and Language 35, 434-453.
Wisniewski, Edward J. 1997. When concepts combine. Psychonomic Bulletin and Review 4, 167-183.
Wisniewski, Edward J., and Bradley C. Love. 1998. Relations versus properties in conceptual combination. Journal of Memory and Language 38, 177-202.
Witkowski, Stanley R., and Cecil H. Brown. 1978. Lexical universals. Annual Review of Anthropology 7, 427-451.
Witkowski, Stanley R., and Cecil H. Brown. 1983. Marking-reversals and cultural importance. Language 59, 569-582.
Witkowski, Stanley R., and Cecil H. Brown. 1985. Climate, clothing and body-part nomenclature. Ethnology 24, 197-214.
Witkowski, Stanley R., Cecil H. Brown, and Paul K. Chase. 1981. Where do tree terms come from? Man 16, 1-14.
Witkowski, Stanley R., and Harold W. Burris. 1981. Societal complexity and lexical growth. Cross-cultural research 16, 143-159.
Wodtko, Dagmar. 2005. Nomen und Nominalisierung im indogermanischen Lexikon. Indogermanische Forschungen 110, 41-85.
Wodtko, Dagmar S., Britta Irslinger, and Carolin Schneider. 2008. Nomina im indogermanischen Lexikon. Heidelberg: Winter.
Wohlgemuth, Jan. 2009. A typology of verbal borrowings. Berlin/New York: DeGruyter.
Wray, Alison, and George W. Grace. 2007. The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. Lingua 117, 543-578.
Wundt, Wilhelm. 1904. Völkerpyschologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte. Vol. 2.2. Leipzig: Wilhelm Engelmann.

Xalilov, Madžid Šaripovič. 1995. Bežtinsko-russkij slovar´. Makhachkala: Dagestanskij naučnyj centr Rossijskoj akademii nauk, Institut jazyka, literatury i isskustva im. G. Cadasy.

Zaefferer, Dietmar. 1991, ed. Semantic universals and universal semantics. Berlin / New York: Foris.

Zalizniak, Anna A. 2008. A catalogue of semantic shifts. Towards a typology of semantic derivation. In: Vanhove (ed.), 217-232.

Zalizniak, Anna A., Maria Bulakh, Dmitrij Ganenkov, Ilya Gruntov, Timur Maisak, and Maxim Russo. 2012. The catalogue of semantic shifts as a database for lexical semantic typology. Linguistics 50, 633-669.

Zamponi, Raoul. 2009. Arawakan: Maipure-Yavitero. In: Lieber and Štekauer (eds.), 584-593.

Zauner, Adolf. 1902. Die Romanischen Namen der Körperteile. Eine onomasiologische Studie. Erlangen: Junge & Sohn.

Zehetmayr, Seb. 1879. Analogisch-vergleichendes Wörterbuch über das Gesammtgebiet der Indogermanischen Sprachen. Auf Grund strenger Etymologie, mit besonderer Berücksichtigung des Lat., Griech., Deutschen, Slavischen und Sanskrit. Leipzig: F.A. Brockhaus.

Summary/Samenvatting

This thesis is a typological study concerned with formal and semantic patterns in the lexicon with a focus on referring ("nominal") expressions. After an introduction to the major research questions in chapter 1, chapter 2 provides a brief discussion of the history of research in this area. In doing so, it shows that interest in the questions has a long but discontinuous history and summarizes literature on the topic that is often not widely known. Chapter 3 describes the general design and outlines the analytical framework of the present study. It is based on a sample of 148 languages all in all for which equivalents to 160 meanings, divided semantically into the domains of nature-related and topological concepts, artifacts, body parts and body liquids, phases of the day and some miscellanea, are investigated. This chapter also develops the classificatory grid to describe the formal make-up of lexical items and the semantic relations between their constituents and/or the various senses they bear, weaving together research strands within linguistics (including historical and cognitive approaches) and Cognitive Psychology. Problems of comparative semantic analysis are amply addressed in doing so. Departing from the basic difference between complex lexical items of the lexical and the derived types developed in this chapter, the following chapter 4 begins with comparative analysis and fleshes out different typological profiles within the lexicon, throughout with reference to the morphosyntactic properties of the languages discussed. This chapter also introduces statistical methods which are frequently applied to a subset of the gathered data in chapter 5, which is concerned with quantitative evaluation of the data and inter alia seeks to provide an explanation for the variation in the number of analyzable lexical items in the lexicon. This is found to correlate with phonological complexity (in particular complexity in the consonant inventory and in the structure of the syllable), as well as with the canonical structure of the nominal lexical root. Chapter 6 moves on to the semantic side of things. Among the topics discussed herein are links between meanings in certain semantic domains as well as common metaphorical transfers across languages. It also addresses the question of both areal patterns in the lexicon as well as globally common strategies to express a particular meaning. A brief outline of possible prospects for similar research in chapter 7 concludes the thesis.

Dit proefschrift is een typologisch onderzoek naar formele en semantische patronen van het lexicon en met name van naamwoorden die naar concrete benoembare zaken verwijzen. Na een introductie van de belangrijkste onderzoeksvragen in hoofdstuk 1 volgt in hoofdstuk 2 een kort historisch overzicht van het bestaande onderzoek binnen dit vakgebied. Het hoofdstuk toont aan dat de wetenschappelijke interesse in dergelijke vragen een lange maar onderbroken geschiedenis heeft, en het geeft een samenvatting van de literatuur over dit onderwerp, die veelal onbekend gebleven is. Hoofdstuk 3 schetst de opzet en het analytisch kader van het onderhavige onderzoek dat gebaseerd is op de equivalenten van 160 betekenissen in een representatieve verzameling van 148 talen. De betekenissen hebben betrekking op de natuur, het landschap, gebruiksvoorwerpen, lichaamsdelen, lichaamssappen, delen van de dag, en nog wat andere termen. Het hoofdstuk beschrijft de algemene opzet van de studie en ontvouwt het schema dat gebruikt wordt voor de classificatie van de formele samenstelling van de lexicale eenheden en de semantische relaties tussen hun constituenten en/of hun verschillende betekenissen, waarbij elementen uit het (historisch en cognitieve) taalwetenschappelijk onderzoek en de cognitieve psychologie worden verweven. Hierbij wordt veelvuldig gerefereerd aan de bestaande problemen binnen de vergelijkende semantische analyse. Hoofdstuk 4 begint met de vergelijkende analyse op basis van het fundamentele verschil tussen lexicale en gederiveerde complexe lexicale eenheden geïntroduceerd in hoofdstuk 3, en brengt verschillende typologische profielen in het lexicon naar voren, waarbij steeds gerefereerd wordt aan de morphosyntactische eigenschappen van de besproken talen. Dit hoofdstuk introduceert ook de statistische methoden die veelvuldig op een deel van de gegevens worden toegepast in hoofdstuk 5, waarin de quantitatieve evaluatie van de data centraal staat, en waarin onder andere geprobeerd wordt een verklaring te vinden voor de variatie in het aantal analyseerbare lexicale eenheden in het lexicon. Het blijkt dat dit correleert met fonologische complexiteit (in het bijzonder met de complexiteit van de consonanteninventaris en de syllabestructuur), als ook met de typische structuur van de (nominale) lexicale wortel. Hoofdstuk 6 belicht de semantische kant van de zaak. Het behandelt onder andere de connecties tussen betekenissen in bepaalde semantische domeinen en veelvoorkomende metaforische overdrachten tussen talen. Ook worden hier vragen aangesproken met betrekking tot areale patronen in het lexicon en wereldwijd veelvoorkomende strategieën om een bepaalde betekenis uit te drukken. Het proefschrift wordt afgesloten in hoofdstuk 7 met een kort overzicht van suggesties voor mogelijk toekomstig onderzoek.

Curriculum Vitae

Matthias Urban was born October 2nd 1982 in Filderstadt and grew up in Tübingen. After attending grammar school, he studied General Linguistics, Comparative Literature, Philosophy, Romance Studies, and Islamic Studies at the Rheinische Friedrich-Wilhelms-Universität Bonn and the Universität zu Köln, earning an M.A. in General Linguistics from the latter in 2008. In January 2009, he joined the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany as a doctoral student.